**Title**
Dismantling the “Black Box:” A Mixed Methods Investigation of STEM Classroom Assessment Experience on Psychosocial Outcomes



**Author**
Chase, Manisha Kaur

**Publication Date**
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Dismantling the “Black Box:”

# A Mixed Methods Investigation of STEM Classroom Assessment Experience on Psychosocial Outcomes

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy in Education

by

Manisha Kaur Chase

2022

ABSTRACT OF THE DISSERTATION

Dismantling the “Black Box:”

Mixed Methods Investigation of STEM Classroom Assessment Experience on Psychosocial Outcomes

by

Manisha Kaur Chase

Doctor of Philosophy in Education

University of California, Los Angeles, 2022

Professor Sandra Graham, Chair

As student-centered classroom practices become more popular, one aspect of the classroom that is increasingly of interest is that of assessment. Lack of student voice in assessment practice has implications for autonomy development, perceptions of power, and motivation, particularly in STEM fields. The current dissertation seeks to bridge the empirical and theoretical domains of power, motivation, and classroom assessment practice in order to realize a more holistic student assessment experience.

Study 1 used qualitative inquiry into classroom assessment experiences of UCLA STEM undergraduate students and professors. Individual interviews ($n$=22), classroom observations, and analysis of classroom syllabi were carried out to understand the lived experience of participants. Coding was conducted on interview transcripts with discourse analysis of field notes and content analysis of syllabi used to corroborate these experiences. Themes included

issues of implicit versus explicit power, foundational need for trust, and motivation in flux. These are discussed as “It Is How It Is” – where participants refer to past/current assessment practice—and how “It Could Be”—where participants envision the future of assessment practice. In general, participants were not satisfied with current assessment practice and provided suggestions for future practice.

Study 2 examined the effects of a classroom intervention that engages student voice in assessment on student perceptions of power, motivation, and attitudes towards assessment. First-year students enrolled in a STEM cluster course ($n$=240) took a baseline survey of measures in Fall with follow-up surveys in Winter and Spring quarters. Half of all discussion sections were randomly assigned to the intervention group in Winter; here, TA’s solicited student voice in participation grading criteria. Linear mixed models were used to analyze effects of the intervention. For all students, perceptions of power increased over time while motivation orientations and grades significantly decreased from Fall to Spring. The intervention had significant impact on first-generation students and those whose TA changed from Fall to Winter.

Findings from the current dissertation have implications for practice and policy. As the landscape of higher education adapts as a result of the COVID-19 global pandemic, it is imperative that assessment practice adapt alongside. Participant perceptions point to a crucial need for equitable classroom assessment that ensures learning and success for all.

*Keywords*: classroom assessment, student voice, power, motivation, critical theory

The dissertation of Manisha Kaur Chase is approved.

Alison Bailey

Kimberley Gomez

Jose Felipe Martinez

Sandra Graham, Committee Chair

University of California, Los Angeles,

2022

# DEDICATION

The contents of this dissertation are dedicated to my parents:

Nirpal Singh Chase & Hardeep Kaur Chase

Drafting this dedication felt as though it took more time than the dissertation itself. I don't know how to adequately put my feelings into words.

You crawled so I could walk. You waded so I could swim. You jumped so I could fly. There are no words to express my gratitude for the countless sacrifices you have made. I am here today and successful because of you. My accomplishments are your accomplishments. These degrees should have your name on them, too. Thank you for making me the person I am. My passion for education, equity, inclusivity, and a fair future for *all*, stems from the lessons you have taught me. I am empathetic, caring, hard-working, and dedicated because of you. These are hard qualities to find and even harder to teach. But they are embedded in me because you showed me their importance.

I know I'm not *quite* the 'doctor' you would have hoped for, but I promise to continue to make you proud. I love you.

**TABLE OF CONTENTS**


INTRODUCTION 1

LITERATURE REVIEW 2

Theoretical Framework 2

*Assessment* 5

*Power* 6

*Motivation* 8

*Assessment and Development* 11

*Current Assessment Practice* 14

CURRENT STUDIES & RESEARCH QUESTIONS 17

Researcher Positionality 18

STUDY 1 19

DATA COLLECTION AND MEASURES 19

PARTICIPANTS AND CONTEXT 24

METHODOLOGY & DATA ANALYSIS 26

FINDINGS & EMBEDDED DISCUSSION 28

It Is How It Is 32

*Logistics* 32

*Hostages* 34

*Micromanaging* 39

*Incongruence* 42

*Efficiency* 45

*Long Term Effects* 49

It Is How It Is: Summary 53

It Could Be 54

*Flexibility* 54

*Liaison* 58
*Positive Experiences* 61
*Active Agent* 64
*Congruence* 67
*Meaning & Purpose* 69
Classroom Observations & Syllabus Analysis 73
Student Suggestions 74
Professor Suggestions 77
STUDY 1 CONCLUSION 80
Link to Study 2 83
STUDY 2 84
DATA COLLECTION & PARTICIPANTS 86
The Intervention 89
OPERATIONAL DEFINITIONS & MEASURES 92
RESULTS 94
Descriptive Statistics 94
Linear Mixed Models 95
Open Ended Responses 97
*Intervention Group Students* 97
*Intervention TA's* 100
DISCUSSION & LIMITATIONS 101
FUTURE DIRECTIONS 105
CONCLUSION 106
APPENDICES 108
REFERENCES 137

## LIST OF FIGURES AND TABLES (ORDER OF MENTION)

**Study 1**

Appendix A: Semi Structured Interview Protocol 109

Appendix B: Semi Structured Observation Protocol 115

Table 1.1: Qualitative Interview Participant Demographics (Chronological Order) 108

Figure 1.1: Most Frequently Used Assessments 30

Figure 1.2: Least Frequently Used Assessments 30

Figure 1.3: Index of Qualitative Findings: Experiences and Corresponding Consequences 32

**Study 2**

Figure 2.1: Graphic Timeline of Intervention 89

Appendix C: Sample Intervention Script & Protocol 116

Appendix D: Sample Student Survey (Fall, T1) 120

Table 2.1: Summary of Variable Means and Standard Deviations over Time by Group 129

Table 2.2: Summary of Bivariate Correlations for All Participants at T1 130

Table 2.3: Linear mixed model with Intervention Group Status and Longitudinal Effects predicting Perception and Performance Variables 131

Table 2.4: Linear mixed model with Intervention Group Status and Longitudinal Effects predicting Motivational Orientations 132

Table 2.5: Linear mixed model with Intervention Group and First Gen Status interaction predicting Performance Approach Orientation 133

Table 2.6: Linear mixed model with Intervention Group Status and TA Match interaction predicting Power 134

Figure 2.2: Differential effect of intervention for first generation students in intervention group vs. control on performance approach orientation 135

Figure 2.3: Differential effect of intervention for students with no TA match versus TA match on perceptions of power 136

Table 2.7: Intervention Improvement Suggestions from Students 100

ACKNOWLEDGEMENTS

Dr. Sandra Graham and the Graham-Juvonen lab for their support over these past few years. No matter how radical or drastically different my research and career interests might have seemed, I always received honest, thoughtful feedback and unwavering support from you all. I sincerely appreciate it.

Special shoutout to Ginny, Diane, Briana, Annie, and Ramon from HDP/HEOC—just a few of the countless cohort peers who have always been so warm, kind, encouraging, and helpful. Whether it was commiserating about graduate school or sharing sources and resources, I truly enjoyed our times in classes together and running into each other in the halls of Moore. Academia can be a lonely place and I am very thankful to have had your support at one time or another.

Can't forget my day ones: Guíge (Mandeep), Tina Penji, Hailey, Amrita, Amber, & Arsh. Guíge—thank you for Pavilions snack runs, advice on millions of miscellaneous things, and feedback on my teaching. Hope we get to 'graduate' together and then I think a 'doggo' is in order to celebrate our accomplishments. Tina Penji—I know (if we're lucky) we only get to see each other every couple of years. But your support, and our regular dose of pange have always kept me going. No matter how far apart, I have always felt and appreciated your sisterly love. Hailey—thank you for always being there with wine and wisdom anytime things went awry. I am forever grateful for our friendship and endless relatable memes. I can't wait to call you 'Doctor' soon! Amrita—thank you for always being real and bringing me much needed humor and perspective. You give me hope for the future of our judicial system. Arsh—thank you for listening to my endless rants: from grad school stress to pandemic politics, teaching mishaps to Punjabi music. You've taught and modeled for me what it looks like to unapologetically pursue one's goals. And your patience and like-minded foodie suggestions (ultra-paneer ftw) have

always been welcomed. Amber—we've sure come a long way since middle school. Thank you for Stonefire catch-up sessions and lending an ear in tough times. From Lawrence Class of 2008: here's to Doctor Graduates of 2022!

There isn't enough room for me to individually thank my extended family; to the Families: Chase, Lotay, Saund, Matharu, Ghataura, Bhamra, Bhari, Pall, and all of the families in between: thank you for your continued blessings, love, and good wishes.

Finally, I must thank Waheguru: ੧ੴ ਸਤਿ ਨਾਮੁ ॥ I would be nowhere without His 'serendipitous' support. Every time I've failed, He has shown me why it was for the better. He has closed doors on me only to open bigger ones shortly after. He has always reminded me: ਸੋ ਸਤਿਗੁਰੁ ਪਿਆਰਾ ਮੇਰੈ ਨਾਲਿ ਹੈ (so satgur pyara mere naal hai). Sikhi is the reason I will always fight against injustice and prejudice, wherever it lives.

VITA

**Manisha Kaur Chase**

EDUCATION

University of California Los Angeles

Master of Arts: Human Development & Psychology (2017-2019)

Institute of Education: University College London

Master of Arts: Educational Assessment (2016-2017)

Bucknell University

Bachelor of Arts, Education (2012-2016)

Bachelor of Arts, Psychology

DISTINCTIONS

- Posse Foundation Full-Tuition Leadership Merit Scholar
- US-UK Fulbright Scholar
- Dissertation Year Fellow, UCLA
- Teaching as Research Scholar, CIRTL

SELECTED CONFERENCE PRESENTATIONS

*It is What It is: Student and Faculty Perceptions of STEM Assessment Practice* presented by invitation at annual Assessment Institute IUPUI Conference (October 2021)

*Student Voice in Classroom Assessment: Considering Autonomy, Power, & Motivation*, Discussion Lead at Association for Assessment of Learning in Higher Education (June 2020)

*Student Perceptions of Motivation & Power as Indicators of Formative Assessment Employment* presented at annual British Education Research Association (BERA) Conference (September 2018)

PUBLICATIONS

Chase, Manisha K. (2020) "Student Voice in STEM Classroom Assessment Practice: A Pilot Intervention." *Research & Practice in Assessment, 15* (2), 1-14.

Chase, Manisha K. (2019) "Critical Assessment: A student-centered approach to assessment." *Papers on Postsecondary Learning and Teaching, 3,* 32-37.

# Dismantling the "Black Box:"

## A Mixed Methods Investigation of STEM Classroom Assessment Experience on Psychosocial Outcomes

"If we were unconstrained by our historic testing legacy, what kind of system might we create, and how would it fit within our schooling processes?" (Stiggins, 2014, p.6). Rick Stiggins poses this provocative question in his book *Revolutionize Assessment* in order to challenge educational stakeholders to think differently about the way assessment has traditionally been carried out in our classrooms. The global COVID-19 pandemic has brought such questions to the fore as practitioners consider the inclusivity of their classrooms (Kinzie, 2020).

The student voice has historically been sidelined in our "testing legacy," demonstrating the disproportionate power dynamics of classroom assessment practice (also known as the "Black Box" of assessment; (Black & Wiliam, 2010). This disproportion has adverse implications for student autonomy development and motivation. Given literature that boasts the effects of autonomy development on student motivation (Chirkov, 2009; Connell & Wellborn, 1991), and given the importance of student motivation on student academic achievement (Graham & Weiner, 1996; Linnenbrink & Pintrich, 2002), one-sided assessment practices may intuitively lead to poorer motivational and academic outcomes. Thus, the study of student involvement (or lack thereof) in assessment practice should consider issues of student autonomy and motivational development and how these may ultimately affect student outcomes.

Research that attempts to bring students into the assessment conversation often focuses on the psychometric properties of the assessments (e.g., their inherent validity or reliability), without evaluating the psychological by-products of students' perceptions of power, autonomy development, and motivation (Cho et al., 2006; Naizer, 1997; Ross, 2006). In contrast, literature

that highlights power differentials in the classroom provides theoretical justification for change but gives little in the way of practical suggestion or concrete empirical findings for how such transformation of assessment discourse affects student success (Evans & Boucher, 2015; Saulnier et al., 2008; Simmons & Page, 2010).

Thus, the current dissertation attempts to reveal these gaps in the literature and provide evidence for the need to explicitly uncover the power disparity present in current assessment practices in higher education. Two studies are then proposed to bridge these gaps: one seeks to provide a qualitative understanding of classroom assessment experience relative to perceptions of power and motivation, while the second involves an empirical investigation of student perceptions of power and motivation given an intervention that explicitly engages students in the assessment of their classroom participation.

## Literature Review

### Theoretical Framework

Critical theory, self-determination theory, and new measurement theory in partnership provide a developmental and contextual foundation for a discussion on power and motivation in the realm of educational assessment.

#### *Critical theory*

Critical theory, cited from sociologist and philosopher Max Horkheimer, is meant to "liberate human beings from the circumstances that enslave them" (Bohman, 2016; Horkheimer, 1982). Critical theorists seek to explore and shift an existing social hierarchy that affords power to some while disadvantaging others. Thus, critical theory requires a perspective embedded in specific socio-political contexts—argued to be the result of any given intersection of history, time, and institutional practice—in order to establish normative practices that "enslave," followed by critique and action in response to such practices. Critical theory

assumes that *all* acts are political; inaction or those that claim to be apolitical simply feed into existing norms and perpetuate a certain hierarchal structure. Thus, critical theory calls for an *explicit* challenge to the status quo. By highlighting those disenfranchised by the status quo, critical theory and its practices move toward the end goal of having people be critical consumers of what has been normalized in their context and be equipped with the knowledge and desire to enact change.

A variety of theoretical lenses can be used to view assessment practice including positivist/scientific, practical/hermeneutic, critical, or post-structuralist (McKellar, 2002). A positivist or scientific lens would advocate steadfast 'Truths,' which often manifest in what we consider traditional exams. In contrast, a critical approach would challenge assessment practices by evaluating to whom and for what reason these practices cater. Ultimately, viewing assessment from a critical perspective requires the larger institution to serve the needs of its student body, as opposed to students conforming to the institution.

Critical theory may help revolutionize the way we think about our current educational assessment practices by highlighting those practices that disproportionately designate power, in order to shift the existing hierarchy and potentially boost student perceptions of motivation and power.

***Self-determination Theory***

Self-determination theory (SDT) is a theory of motivation that distinguishes motivation as being autonomous versus controlled (Deci & Ryan, 2008; Deci & Ryan, 1985). Autonomous or intrinsic motivation occurs when an individual experiences volition, through internal factors free from outside pressure or reward, versus controlled or extrinsic motivation in which an individual's actions are a result of external rewards or punishments.

In education, SDT argues that student motivation to learn can be explained by how well teachers encourage individual growth– including via autonomy (Connell & Wellborn, 1991; Skinner & Belmont, 1993). Students who cite their classroom as providing autonomy support have been shown to have better academic outcomes (Black & Deci, 2000), while controlled motivation has been demonstrated to have a negative effect on student academic achievement (Manganelli et al., 2019). When an educational environment is perceived as having no room for control, self-determination—and consequently, motivation—experience decline (Deci et al., 1989). In this way, SDT has been aligned with critical theories in that its focus is on the larger system or context which enables or inhibits individual autonomy (Ryan & Niemiec, 2009). With its focus on agency and autonomy, SDT shifts the conception of teacher as the sole power-wielding authority in a classroom context in the same way critical theory demands an examination and dismantling of hierarchies in a given context.

### *New Measurement Theory*

In order to scaffold the focus on classroom assessment practice, new measurement theory is used here as a lens through which assessment is conceptualized (Bonner, 2013). While more traditional assessment and measurement theory (Traub, 2005; van der Linden & Hambleton, 2013) tend to focus on assessment practice in a silo of its inherent qualities, new measurement theory grounds assessment practice in the "interpretations" of assessment score meaning by stakeholders (including students). This social-constructivist view—now the more common assessment perspective—suggests that assessment, judgements made in its regard, and subsequent uses are centered in context rather than having a predetermined and "fixed" meaning. The acknowledgment and grounding of assessment theory in social context inevitably gives rise to concerns of equity—including power, and which voices are included in the meaning-making

of assessment use. In this way, new measurement theory illuminates the periphery of assessment practice, which, from a critical perspective, must be acknowledged towards understanding and acting upon existing normative practice.

**Assessment**

Assessment in the context of education is formally defined as the "process of gathering evidence, and secondly interpreting that evidence in the light of some defined criterion in order to form a judgment" (Broadfoot, 2007). The aforementioned "evidence" of learning can manifest in a variety of ways; while the most common conception of such evidence is test scores, this is just a fraction of the assessment evidence produced within a classroom. Synchronous verbal answers during class, essay assignments, and portfolios are just a few examples of additional evidence that can be used toward forming judgement about learning. This feedback serves two purposes: it can be used not only to inform instructor practice, but also to provide students with an anchor of what and how to improve or progress.

An ongoing dialogue in practical and theoretical assessment considerations is that of purpose and product. Summative assessment is often aligned with those assessments that summarize learning up to a certain point (i.e., following a unit or semester), whereas formative assessment is construed as being focused on providing feedback for the purpose of enhancing student learning (Brookhart, 2004). In the literature there has been much debate about whether summative and formative assessments are distinctly different from one another (Stobart, 2008), or rather, simply an extension of one another (Newton, 2007). More specifically, the debate has centered around whether formative or summative practices cater to assessing as a process *or* a product. In this regard, Taras (2008) argues each practice can take the form of *both* process and product. While this dissertation will tease apart differences in summative and formative assessments as they appear in both the literature and impending studies, stakeholder perception

of assessment purpose and product will overshadow this terminology and distinctions. I focus on the way instructors and students practically perceive and construe assessment (including if they use formative and summative descriptors), rather than the technical classification of the assessments themselves.

Regardless of the alleged type of assessment employed, purpose and determination of purpose are important facets of assessment practice as they relate to validity of assessment for those set out purposes. How purpose is conceptualized and who has a say in its conceptualization (i.e., where power lies), is a key concern of this dissertation that I will address in later sections.

**Power**

In the education context it is necessary to consider identity in relation to learning; more specifically, how are identity enacted in the classroom on an individual and social level (Pryor & Crossouard, 2008). Power is conceptually defined here as "student voice" (Bain, 2010) and "potential for influence" (Hosek & Houser, 2018) within a classroom. The traditional view of education implies student subordination to their teachers—a notion that is at times still in place (Menges, 1977). Given the intersections of student and teacher identities within the classroom, it is fair to say that power may not be held equally by every individual.

Assessment, in particular, is one sphere of education where identity and power are at odds. Assessments that affect decisions made about a student are often developed without student counsel (Boud, 2007). Moreover, it is assumed that only teachers have the esoteric knowledge or expert power (French et al., 1959) necessary to create effective assessments. In this respect, student assessment history involves a regime of truth where assessment is dominated by teachers (Asghar, 2012; Foucalt, 1977). This, then, creates a hierarchy that perpetuates student dependency on teacher judgement (e.g., grades) (McCroskey & Richmond,

1983) which may stifle student autonomy (Ryan & Niemiec, 2009) and exaggerate the power stratification (Sadler, 1989).

While assessments with formative purpose are intended to provide students with more agency and power in the classroom (Pryor & Crossouard, 2008), it has been argued that a lack of theoretical basis (Taras, 2012; Taras & Davies, 2017) has led to confusion, distorted use, and therefore disparate outcomes in practice for both instructors and students (Taras, 2008). Consequently, if students are not able to discern the possibility for power within classroom assessment practice (e.g., through conflation of feedback and grades in the example of formative misuse [Taras & Davies, 2013]), this may lead them to perceive a lack of power in the classroom. In order to reap the benefits of empowering students, it becomes necessary to understand the power dynamics in a classroom—including how they are perceived, produced, and sustained—in attempts to engage students in the learning process.

Given the intangible nature of power, studies of power in the classroom often rely on *perceptions* of power. In an ethnographic exploration of classroom power, Sidky (2017) used discourse and conversational analysis of video-taped classroom interactions and interviews to understand student-teacher power relations in a graduate classroom. Findings were summarized as a negotiation of coercion and consent. Coercion typically originated from teachers (often subtly disguised), while consent stemmed from students' belief that their instructors were the authority in their classroom, and that they should "surrender as part of an unwritten contract" (p. 190). Institutional expectations were cited as one potential cause. In one example, a teacher explained they felt they would be judged by administration and faculty peers if they allowed students to have power in the classroom. This fear led to an inability to relinquish power in their classroom. Sidky argues that "explicitness," is the key to encouraging

student power and teacher relinquishment of power. This explicitness comes from an acknowledgement of the ideologies that underlie classroom practices and calls for the need to re-evaluate attitudes towards power dynamics in the classroom.

In attempts to negotiate power in the classroom, dialogue has been found to be an effective tool in the current literature (McLean, 2018). In a case study of Australian university professors, researchers found dialogue to be necessary to promote feelings of trust and power felt by both professors and students in the classroom. For example, through conversations and feedback, students expressed feeling validated and welcome in their academic circles, which subsequently prompted a desire to achieve more and work harder. In this way, dialogue may be a necessary strategy for challenging the current state of student power in assessment practice.

In all, the current research on power in the classroom suggests that power is still predominantly held by instructors. Dialogue, however, appears to have promise for reforming power differentials. That being said, much of the work on power is qualitative in nature and does not often link to other psychosocial or academic outcomes such as motivation or grades. Finally, this work is contextualized in the classroom as a whole, but rarely in relation to the assessment domain specifically, where power is arguably most stratified.

**Motivation: Intrinsic versus Extrinsic**

Motivation in the realm of schooling is concerned with drive, attribution, and goals that determine the type and extent of action taken by an individual. This social-cognitive approach focuses on how individuals conceptualize classroom situations and consequently react to them (Dweck, 1986). Goal-oriented theories of motivation (i.e., performance versus mastery), focus on *how* effort is directed toward certain outcomes but arguably neglect *why* such outcomes are sought after (Deci et al., 1991). The *why* is addressed in self-determination theory which rests on the idea of competence and individual need.

In a meta-analysis of 153 studies across a variety of domains including school and work, Cerasoli and Nicklin (2014) set out to understand how intrinsic and extrinsic motivation have predicted performance over 40 years of research in this field. Authors concluded that intrinsic motivation explained more variance in the *quality* of work, while extrinsic motivation explained more of the variance in *quantity* of work. One suggestion for practice proposed linking tasks that require mundane repetition or increased quantity of production to extrinsic rewards, while that which necessitates "personal investment, complexity, and overall quality" (intrinsic) should be less tied to such rewards (p. 21). In the context of education, it is argued the latter case is most applicable. However, given our regime of truth which includes grades, ranking, and level progression, a complete avoidance of external rewards is strictly speaking: impossible (Foucalt, 1977).

Given the contextual dimension of motivational adaptivity, it should be noted that motivational orientation cannot be simply dichotomized. The Motivated Strategies for Learning Questionnaire, or MSLQ (Pintrich, 1991), is a measure often used in undergraduate populations which gauges, among other things, intrinsic and extrinsic motivation in the classroom. For example, one item in each touches on what would make a student "most satisfied" in their class; the intrinsic scale cites "trying to learn as much as possible" while the extrinsic scale cites "getting a good grade." Logic suggests the plausibility of simply altering one of the subscales to represent the "opposite" motivational orientation (i.e., reversing the scale of extrinsic items and adding them to a participant's intrinsic score) in order to sort participants into being intrinsically versus extrinsically motivated. However, a recent Rasch-based construct validity study on the MSLQ revealed that it is actually a matter of the *degree* to which participants are intrinsically or extrinsically motivated (Nielsen, 2018). Participants, though, typically scored higher on one and

lower on the other such that there is generally an individual "preference" for one motivational orientation over the other. While this study urges a shift away from the strict divide of intrinsic *versus* extrinsic orientations, it does provide evidence for having a greater inclination towards one or the other.

Further, recent literature suggests that these motivational inclinations are not permanent: having a higher degree of intrinsic versus extrinsic motivation is not a stable characteristic. In a longitudinal cluster analysis of student motivational profiles, Hayenga and Corpus (2010) classified students according to their degree of intrinsic and extrinsic motivation. Students were classified into one of four categories: as having high or low quantity plus high or low quality of extrinsic and intrinsic motivation. From Fall to Spring semester, researchers found 43% of students had changed motivation cluster membership. More specifically, the general trend indicated students moved towards poor quality and low quantity memberships (e.g., less intrinsic motivation). This change is cited as a potential result of the "escalating pace" of the classroom, in addition to "classroom exams" that become more prominent towards the end of the academic year. This study suggests that motivational orientation is not resistant to change, and in fact, may be explained by contextual factors such as perceptions of the classroom or assessment practices.

The intricacies of theory and semantics aside, motivation in general is hailed as a critical mediator of student achievement in the academic context (Covington, 2000; Kusurkar et al., 2013; Sharma & Sharma, 2018). It is important to note that like many socioemotional factors, motivation is dynamic and *context* specific (Linnenbrink & Pintrich, 2002). In this way, it is necessary to understand the classroom contexts, such as instructional methods and assessment practices, that help or hinder the development and maintenance of motivation for students.

Student power and motivation will herein be discussed in conjunction with the theoretical framework of the current paper, with particular attention to the developmental importance of these in relation to classroom assessment practice.

**Assessment and Development**

Prior to assessment and student voice being formally combined for any student population, it is necessary to understand how student development relates to assessment. From a psychometric point of view, it may be argued that development has no place in testing or assessment outcomes. However, a contention for why such a discussion is needed is as follows.

Validity of assessment purpose—a key concern in assessment production —may be heavily impacted by student development. Validity is defined as the extent to which assessment purpose and assessment output align (Harlen, 2005). The importance of considering assessment purpose is reiterated here, as validity is always reported in regard to the 'validity of the purpose' of an assessment rather than being an inherent feature of the assessment tool itself. In fact, Crooks, Kane, and Cohen (1996) argue that validity "is the most important consideration" in assessment. However, assessments are often prone to "pollution" (Haas et al., 1990) of validity, otherwise known as construct-irrelevant variance (Haladyna & Downing, 2005). These may include 'teaching to the test' or resource availability, for example. This idea is expanded in Messick's (1998) consequential validity which suggests that the interpretation of assessment outcomes have social consequences which are "fundamentally contributory to…meaning and hence to construct validity" (p. 41). An often overlooked example of consequence to the validity of assessment purpose, is student motivation (AERA et al., 2014).

If students are not motivated to apply themselves, this may affect the effort they put forth, their perception of the gravity of an assessment, and ultimately, their performance (Wise, 2020).

This distorted performance is a major concern for those interested in assessments valid for various purposes, as it leads to misinterpretations of students' actual learning and thereby potential misuse when acting upon this misinterpretation (Paris et al., 1991). Thus, development (in this case, of motivation) is a necessary consideration in assessment conception.

Linking back to motivational orientations, it is safe to say that assessment practices in which students are the receivers rather than co-creators may lead to a perception of stifled student autonomy, and thus, inhibited autonomous motivation. Until now, assessment has often been associated with "judgment," and less so with "participation" (Boud & Falchikov, 2006). This has led to a push for assessment practices which acknowledge student personal and academic development (Knight & Yorke, 2003, p. 477).

In general, researchers find decreasing student motivation towards assessment practices as students progress from K-12 (Paris et al., 1991). While this lack of motivation was reported in response to standardized testing, it can potentially be related to other forms of assessment and may continue as students enter a university. Bringing students into the classroom conversation has been cited as increasing motivation and perceptions of power (Astin, 1999; McLean, 2018). Therefore, it is argued that this same concept of bringing students into the conversation can be applied specifically in the realm of educational assessment, to bolster student autonomy development and thereby potentially improve the validity of purpose of our assessment practices.

### *Higher Education & STEM Context*

While the hope would be to bring students into the conversation at *all* levels of educational assessment, it is posited here that universities and students in higher education are the ideal starting point. Universities allow more agency and flexibility when it comes to assessing in the classroom (Naidu, 2017).

Moreover, metacognition and critical thinking have been shown to reach a pinnacle in late adolescence which coincides with typical undergraduate years (Dwyer, 2017; Palmer et al., 2014). Having the ability to understand one's own learning and apply evidence is important in any educational setting. This is especially so with assessment, where assessment uses may have significant consequences. In fact, most applications of engaging students in assessments take the form of self-assessment or peer-assessment, which involve college student populations (Ashenafi, 2017; Falchikov & Boud, 1989; Wanner & Palmer, 2018). Thus, a meaningful endeavor to involve students in assessment practice would be most appropriate with a population that currently possesses such skills and has been established in the literature as a developmentally appropriate population to collaborate with.

Another important contextual intersection for the proposed studies is STEM fields in higher education. The areas of science, technology, engineering and mathematics (STEM) in education are known to be "cutthroat" disciplines (McGee, 2016) with "individualistic weed-out culture" (Daily et al., 2007) that are largely grounded in White, middle-class, masculine norms (Fabert et al., 2011). Retention in STEM is low across the board but known to disparately affect ethnic and gender minoritized groups (Chen & Soldner, 2013).

At the heart of many explanatory factors that attempt to narrow down the source of this phenomenon (including a lack of belongingness, loneliness, and feelings of helplessness) are assessment practices and grades which not only neglect to include these marginalized voices in their conception, but also expose the "gatekeeping" function of these assessments (and therefore the gatekeeping function of STEM as a whole) [National Academies of Sciences, Engineering, and Medicine, 2016]. These feelings of distance between students' respective identities and STEM content are reflected in assessment practices that are not only traditional in nature

(quizzes and tests resting heavily on memorization techniques), but also "devoid of deep connection" to real life given their focus on isolated facts (Martin-Hansen, 2018; Momsen et al., 2010). Therefore, when considering assessments that serve to empower students, STEM is a context that may especially benefit from such intervention, but also one that that requires a unique consideration of its culture: including hegemonic norms and the way those norms are perceived and acted upon by both instructors and students.

**Current Assessment Practice**

The current literature review has provided conceptual definitions of educational assessment, motivation, and power, as well as established a theoretical scaffold and population of interest. This final section will bridge these areas of research by citing current practice and discourse that combine assessment and student voice.

Practice that has attempted to bridge student voice and assessment has often been confined to the existing framework of power. For example, professors may have students engage in varied assessment types or ensure that students have practice exams that align with actual exams (Saulnier et al., 2008; Turner, 2014). While these practices have student welfare in mind and attempt to cater to issues of fairness within assessment, they do nothing to address the imbalance of decision-making power in classroom assessment. That is, these assessments are pre-determined and do not consider student voice in assessment development.

With special attention to university populations, it is argued that current assessment practices fail to prepare students for "a lifetime of assessing their own learning," which they need as they move into the workforce (Boud & Falchikov, 2006, p. 400). Appropriate assessment practice, researchers argue, should provide students the skills with which they may appraise and make judgment about both their own and their fellow peers' work, towards the end of taking appropriate action in any number of unknown future contexts (Boud & Falchikov, 2006).

Students are often left out of the conception of assessments used to make judgements about their learning, and thus, left out of the process of learning *how* to assess. Thus, student participation in assessment is not only necessary for the purpose of bolstering motivation to learn, but also for the purpose of preparing students for a future beyond the classroom.

In one example of students involved in "flexible assessments," a sample of undergraduate accounting students was given the opportunity to assign their own weightings to the existing assessments in attempts to gauge what effects, if any, this participation would have on student attitudes, motivation, and grades (Pacharn et al., 2013). Researchers found that while the timing of determining what their weightings would be (e.g., earlier in the semester rather than towards the end) did not have a significant effect on any of the variables of interest, those who were allowed full-flexibility both in terms of weighting and the timing of determining their weighting found participation in assessment as useful, cited increased motivation, and had overall better grades than both the control group and the group in which students had to set their weightings early on in the semester. While this study's interpretations may be limited by the fact that students would likely assign higher weight to assessments in which they performed better in (hence better outcomes), results do point to the possibilities for improving student motivation as a result of involving them in assessment practice.

Current research demonstrates some student involvement in various assessment practices. However, these studies have faltered in supporting true student autonomy and participation in the following ways. While students have been given the opportunity to choose a topic they enjoyed or allowed to determine assessment weighting, they did so within the confines of an existing assessment format. In other words, students were not involved in the discourse on what purposes the assessments would serve, or which assessments might best serve said purposes. Students

were involved only *after* these decisions had been finalized. Moreover, while studies approach student participation in assessment from theoretical perspectives that encourage student autonomy and student-centered learning, they fail to explicitly acknowledge issues of power and bridge such theory (e.g., critical theory) to assessment practice. Finally, it should be noted that the studies in this area rarely, if at all, probe students' perceptions of power as a result of being involved in assessment dialogue.

In one example of an explicit discussion of assessment and critical theory, Bain (2010) calls for assessment that "empowers" students, and specifically notes critical theory as a necessary theoretical underpinning toward fostering such empowerment. The author provides a clear review of assessment literature and how critical theory may scaffold a reinterpretation of power in the assessment context. However, it does not lead to an empirical proposition for studying or implementing such practice. Moreover, this work is proposed in the context of UK higher education, which differs not only by shallow indicators such as time to degree (traditionally three in UK versus four in the US), but also in respect to assessment culture. For example, UK universities are all subject to periodic and independent review by the Quality Assurance Agency for Higher Education (QAA) for their assessment standards and practices, while there exists no such central organization equivalent in the US (QAA, 2019). In attempting to link assessment practice and critical theory, it is necessary to specify the particular socio-political context of study, as this context brings its respective status quo, power differentials, and bureaucratic limitations. Thus, it is argued here that an empirical application of these ideas to the US higher education context, with an added focus on student psychosocial and academic outcomes, is necessary. It is here—between the camps of student

involvement in assessment practice, considerations of power and critical theory, and the development of autonomy and motivation—that the current dissertation resides.

## Current Studies and Research Questions

Hearkening back to Stiggins' (2014) call for 'revolutionizing assessment' practices, it is imperative that we in the educational community "embrace a new role for students" (Stiggins, 2017). Students' historic lack of voice and argued lack of power in the realm of education may have downstream effects on academic achievement. While there exists a body of literature that pushes for change in classroom assessment toward a more student-centered model, much of it is theoretical in nature and does little to provide tangible illustrations for practice. When student voice has been considered in assessment research it often ignores issues of power and by doing so, rarely involves students in a meaningful way. Moreover, few studies have used quantitative or mixed methodology in order to evaluate the effects of student voice in assessment practice. Finally, the psychosocial by-products of inviting student voice into the assessment discourse such as perceptions of power and motivation, have largely been overlooked.

To address the void in research, the current pair of studies in the area of assessment and student voice explicitly set out to understand power dynamics of current assessment practice, engage student voice in meaningful ways (i.e., during the assessment production phase), employ more quantitative methodology, and most importantly, explore the potential effects of involving students such as changes in perceptions of power and motivation which may shed light on eventual academic outcomes. This dissertation begins with a qualitative understanding of undergraduate assessment experience at UCLA, followed by an intervention engaging students in the assessment process which quantitatively gauges change in perceptions of

power, motivation, and attitudes towards assessment. It is well documented that STEM fields, particularly, report staggering dropout rates, lack of student motivation, and grading disparities (all of which are exacerbated in marginalized populations) [Lowery, 2010; Smith, 2019; Van Soom & Donche, 2014]. Thus, the current studies seek to explore these aforementioned assessment and student voice concerns in the specific context of STEM classrooms. The research questions address:

1. What is the current classroom assessment experience of undergraduate students and professors at UCLA, particularly in regard to perceived power and motivation?
2. What are the effects of an intervention that meaningfully involves student voice in assessment practice on student perceptions of power, motivation, attitudes towards assessment, and academic achievement in the context of a first-year STEM course? Moreover, how do these perceptions change over time?
3. Does student perception of power mediate the effect of motivational orientation on academic achievement in a first-year STEM course? How does this effect vary as a result of incorporating student voice into assessment practice?

More specifically, in Study 1 I use qualitative methodology to investigate and interpret participant experience in response to the first research question. In Study 2, with an experimental approach, I use a novel intervention in a first-year STEM course towards answering the remaining research questions.

**Researcher Positionality**

I want to address my positionality and motivation for this research; more specifically my academic position, my assessment privilege, and my future academic aspirations. Majority of my academic experience has been as a student, thus, I am more easily able to empathize with the

student perspective. Moreover, while I criticize the status quo of assessment practice, I have clearly benefited from it as evidenced by my association with the various organizations (such as the Posse Foundation, Fulbright Association, and graduate school at UCLA). Thus, while I may have not been a "straight-A" student, I would like to acknowledge that the status quo has privileged and sided with me as a result of gaining access to these institutions and their support. I could not have gained access to these programs without some sort of evaluative success, so I have been and continue to be privileged by the current ways of doing. Finally, I have a long-term goal and desire to teach. In this way, my methods and interests invariably tend towards ideas and conclusions that are practical and might be easily used by practitioners in the classroom. I encourage you to keep this background in mind as I delve into my research, it's methods, and my conclusions.

## Study #1

### Data Collection and Measures

In this first study, my primary aim was to understand the classroom assessment experiences of STEM undergraduate students and professors at UCLA. Students are often shut out of the discourse of assessment practices that directly influence them. Simultaneously, however, it is unclear whether students even want to be involved in the assessment process in the first place. Moreover, given that the key for change in any assessment practice would be made by instructors, it was equally important that I understand their assessment experiences, too. Thus, my purpose in this study was to understand how students and professors perceive the assessment practices currently employed in classrooms (particularly in regard to perceptions of power and motivation), how they feel about those experiences, and whether they would like to see change, if at all. My rationale is such that by understanding the *current* experience of STEM classroom assessment practice at UCLA there creates potential for the *future* of improving such practice. As

my aim was to understand the lived experiences of students and professors, I employed qualitative methodology.

Qualitative research is concerned with "understanding the meaning people have constructed [and]… how people make sense of their world and the experiences they have in the world" (Merriam, 2009). In order to understand people's actions, it first becomes necessary to understand the meaning people make of their world and experiences. Thus, this method was deemed appropriate for answering the first study question. More specifically, a grounded theory-aligned approach was employed in this study. In the strictest terms, grounded theory is an inductive approach that allows participant data to form theoretical meaning as opposed to having pre-disposed notions of what participant experience might be (Glaser & Strauss, 1967), considered appropriate when not much is known about a subject area (Chun Tie et al., 2019). While assessment experience and issues of power and motivation are not unknown in the literature, much of what we know derives from quantitative research methodology, and rarely combines these areas of study. Given the specific research query of this first study, a grounded theory variation is appropriate here as it seeks to "make knowledge claims about how individuals interpret reality"—in this case, the reality of their assessment experiences (Suddaby, 2006). Thus, "substantive theory" is used to provide a foundation for the grounded theory approach in this study (Glaser & Strauss, 1967). I collected three types of data under this approach: participant interviews, classroom observations, and classroom syllabi.

### *Participant Interviews*

Participant interviews were the primary source of data in Study 1. I conducted individual interviews with 22 undergraduate students and professors in STEM fields at UCLA. Inclusion criteria for the current study were as follows: being a current undergraduate student or professor

at UCLA, 18 years of age or older, conversational in English for the purpose of interviews, and having had a STEM field association. Interviews were conducted in Winter (in person) and Spring Quarter 2020 (over Zoom due to COVID and online instruction) and lasted between 45-90 minutes. Sampling began with email recruitment of convenience (from professional development email list-servs) which then led to snowball sampling and self-referral of participants who were interested in sharing their assessment experience. An audio recording device was used to capture interview conversations, as well as for the purpose of creating written transcripts. I manually transcribed the audio files in full, and both audio files and transcriptions were securely stored on an encrypted device.

***Interview Protocol***

A semi-structured interview protocol relative to students and professors respectively was created and used for these conversations (see Appendix A for full protocol). The protocol probes retrospective experiences with classroom assessment at UCLA, affective responses to assessment experience, concerns, and choice regarding assessment, in addition to perceptions of power and motivation in the classroom. Sample student questions included:

1. When you enter a class on the first day, what are your biggest concerns or questions about the upcoming quarter?
2. Given this definition [of assessment], take a moment to think about all of your classroom assessment experiences here at UCLA. Can you describe a few of those moments that stick out the most in your memory? Why do you think those instances in particular stuck out for you?
3. In thinking about the ways you have been assessed at UCLA, what are the most frequent ways your learning has been assessed? The least frequent?
4. What has been your most enjoyable assessment experience and why?
5. When you think about being assessed in the classroom, how do you feel? What emotions do you associate with various assessment?
6. Imagine you are catching up with a classmate after class. What are some of the kinds of conversations you would be having about that class, particularly regarding assessment?

7. Have you ever been asked to participate in assessment practice? Whether that means coming up with the way you will be assessed, having a say in the assessment process, having flexibility or choice, etc. If so, describe that instance.
8. Switching gears just slightly, what does it mean to you to have power in the classroom? What ways have you felt you have had power in the classroom before? How does it feel to be in a classroom where you "have power" vs. where you feel you don't?
9. Can you describe the extent to which you feel you have power when it comes to classroom assessment?
10. Think about the type of assessments you have had in the past, and now think about your own motivation. How do these assessments factor into your motivation in the classroom, if at all?
11. If your professors could anonymously get to know one thing about your assessment experience, what would that be and why? In a similar vein, if you were given the chance to have a say in the way you were assessed in class, what would that look like? What would you suggest and why?
12. If you were asked to engage in assessment, is that something you would want to engage in? Why or why not?

The professor protocol was similar to that of students' with the substitution of appropriate context (i.e., "Imagine you are catching up with a "*colleague*" vs. "classmate"). Additionally, all professor interviews began with background questions probing how many years the professor had been at UCLA, courses they have taught, and typical student populations in those courses (i.e., transfer students, upper division, etc.).

***Classroom Observation and Syllabi Analysis***

While participant interviews served as the primary source of data in Study 1, a second method for understanding participant experience was using classroom observations in STEM classrooms at UCLA[1]. Observations spanned the duration of a class period (75-110 minutes), and each classroom was visited at least three times throughout the quarter. A semi-structured classroom observation protocol (Brighton et al., 2007) was adapted for the context and purpose of the current study (see Appendix B). This protocol probes the implicit and explicit curriculum

[1] All observations were conducted in-person, prior to the COVID-19 global pandemic

of a given classroom and provides context to additional observations related to assessment dialogue in the classroom. Field notes were recorded both during and directly after each site visit in order to capture, in as much detail as possible, the workings of the classroom and the interactions relative to assessment practice. Observational data were collected from Sites A and B (described below) during Winter Quarter 2020.

Two classrooms were selected as observation sites for their diversity of class size, subject matter, and instructional style. The first, Site A, was a large upper-division Engineering, 110-minute lecture of approximately 100 students which took place in an auditorium-style classroom (capacity for approximately 300 persons). This course was Bruin-casted (video recorded) and required the professor to wear a microphone in order to be heard by students in the class, as well as for the video recording. The professor typically used PowerPoint (Microsoft Office, 2015) to administer lecture material with the help of an electronic pen and a tablet which was used to add 'live' notes to slides.

The second observation site, Site B, was a medium-sized, 75-minute, Physical Science upper-division course of approximately 60 students. This lecture took place in a smaller auditorium-style classroom (in comparison to the first site), with capacity for approximately 100 persons. This smaller classroom size did not require the professor to use a microphone and the classroom was not video recorded for subsequent viewing online. The professor typically used PowerPoint to administer lecture material as well as the *iClicker* electronic response system to poll students throughout lecture (*iClicker*, 2020).

Discourse analysis was used to parse conversations that took place during classroom observations. This type of analysis serves as an "intervention of social and cultural norms, which determines roles and relationships of the participants" (Suciu, 2019, p.2). Given that power is a

phenomenon that exists within the relational ties of individuals, and of key interest in this study, discourse analysis was necessary to uncover the ways in which discourse in the classroom contributes to or shapes the realities of power for participants.

Finally, content analysis was conducted on the syllabi of the classrooms that were observed in this study. Among its many aims, content analysis serves as a method of going "outside the immediately observable physical vehicles of communication and relying on their symbolic qualities …thus rendering the (unobserved) context of data analyzable" (Krippendorff, 1989, p. 403). As the current study seeks to understand power dynamics that may be a result of social/cultural norms, it is necessary to explore the effects of language that is symbolically used to uphold these norms. The course syllabus is typically the main format in which policies and details regarding classroom assessment are recorded, and thus serves as a "boundary object" that embodies and dictates classroom assessment practice (Winget, 2008).

Additionally, throughout the research process I kept a researcher journal that began in Spring Quarter 2019 for the purpose of recording research decisions, raising potential concerns, and maintaining researcher reflexivity.

**Participants and Context**

There were 22 total participants in Study 1—13 undergraduate students and 9 professors, 15 of whom self-identified as females and 7 self-identified males (see Table 1.1). Ten participants self-identified their ethnicity as White, 8 participants self-identified as South Asian or Indian, 2 as multiethnic, 1 as Middle Eastern, and 1 as South-East Asian. At the time of interviews four students were seniors, eight were juniors, and one was a sophomore. Thus, student participants had *at least* 4-10 academic quarters of classroom assessment experience at

UCLA to retrospectively speak of during our conversations. Similarly, professor participants had at least 2-12 years of classroom assessment experience at UCLA.

Of the 10 interviews in Winter 2020, all took place in-person either in private study rooms in the library or in private respective offices. Private study rooms typically consisted of one circular table with 2-4 rolling chairs around it, a large mobile whiteboard in the back (sometimes with study notes from the prior reservation) surrounded by blank walls. Each study room had an entrance of a glass door and wall; thus, passerby's could vaguely see individuals in the room but could not hear the contents of our discussion. Each private office was either an office in isolation with its own entrance, or an office within a cluster of offices that had a main entrance and then an entrance to the respective private office. In most cases, the participant closed the door, but even for those who did not, interviews were never interrupted. As a result of the switch to online instruction due to the global COVID-19 pandemic, the remaining 12 interviews took place over Zoom in Spring 2020 (Zoom Video Communications, Inc., 2020). Almost all interviews had strong connection such that recordings and the transcripts were unaffected by the online format; however, for one interview the participant did have to turn their camera off in order to allow a more seamless audio stream.

In terms of disciplinary representation, 45% of participants were from the Psychology department, 14% from Biology-related subfields, 23% from other STEM fields such as Psychobiology and Physics, with the remaining 18% having experience in both STEM and other fields. Three participants represented dual-disciplines (2 students and 1 professor)—two of them (1 student and 1 professor) represented Psychology plus another STEM field, while the remaining student represented Atmospheric and Oceanic Science plus a Humanities field. Finally, one student came from a Biology-related subfield but recently switched to a Humanities

discipline. While the student does not currently represent a STEM major, their prior experiences with classroom assessment while in STEM were of value and still applied to the current research question, thus, their responses were still included in study findings.

**Methodology and Data Analysis**

I manually transcribed individual interview recordings using the audio recorded files from each interview. These transcriptions were then organized and analyzed within the Dedoose web application software (*Dedoose*, 2018). More specifically, transcriptions were analyzed using both an inductive and deductive coding method (Strauss & Corbin, 1990) or blended approach (Graebner et al., 2012). It should be noted that while the methods here sought guidance from grounded theory, there was a literature review conducted on theories that might inform the experiences of participants (critical theory, self-determination theory, and new measurement theory); thus, the following methods and analysis should not be interpreted as strict grounded theory, as this literature review and existing theory may have affected the inductive round of coding (though not intentioned).

Inductive coding is typically undertaken in qualitative research in order to let participants' experiences speak for themselves. Thus, the first round of coding participant interviews, observation notes, and syllabi were conducted using open coding: a general survey of all responses for initial similarities or themes. During this first round, Lofland et al.'s (2006) aspects of coding including cognitive aspects/meanings, emotional aspects/feelings, and hierarchical aspects/inequalities were used to help derive codes. This first round included seven readings of each transcript and resulted in 10 preliminary codes.

Given the prior knowledge of assessment and issues of power and motivation in the literature, a deductive round of coding was also deemed appropriate for this study (Skjott

Linneberg & Korsgaard, 2019). Thus, the second round of coding was deductive in nature and relied on a pre-determined codebook based on the guiding theories for the current research question. Critical theory lent itself to codes of coercion vs. consent and expertise, Self-determination theory lent itself to autonomous and controlled motivation, and finally, New measurement theory lent itself to meaning of assessment and summative vs. formative codes. This second round included three separate readings of each transcript, each using a different theoretical lens and their six total, respective codes.

I then conducted a third round of coding using axial coding to understand how the initial themes derived from both inductive and deductive processes might relate to one another. I used the code co-occurrence matrix in *Dedoose* to enable this axial process. The matrix provides information regarding the frequencies of code pairings overlapping for a similar excerpt and allowed codes from the first and second round of coding to be compared simultaneously. For the purpose of this round, each pair of codes with a value of 16 or greater (maximum of 79) were highlighted for further review. The excerpts that were similarly coded were categorically evaluated as to whether the codes pointed to the same phenomenon or rather, were related as subsets of one another, or simply had distinct meaning. As recommended by Saldaña (2013), each co-occurrence was documented as analytic memos in my researcher journal with accompanying reasons as to why the codes were combined or left as unique identifiers. As a result of this process, four different pairs of codes were combined. This left a semi-final list of 11 codes.

The final round of coding involved looking for exceptions to the codes and respective themes that had since been arrived at. These were noted in the researcher journal as important exceptions to the rule. This four-round coding process revealed three overarching themes of

power, trust, and motivation which will be discussed in detail in the following section. While a second rater was not secured for this study (for lack of resources), validity and reliability (trustworthiness) of the current study were ensured through triangulation of multiple data sources (participant experience via interviews, classroom experience via observations/discourse analysis, and "boundary objects" via content analysis). Moreover, member checks with participants were conducted upon the first draft of findings in order to ensure that data (specifically, participant direct quotes and descriptions) were accurately reported and in such a way that participants approved for public dissemination.

**Findings and Embedded Discussion**

Prior to the researcher defining classroom assessment and providing examples of assessment in practice, all student interviews began with the question of what concerns students typically have on the first day upon entering a new class. This question was asked to gauge whether and to what extent assessment was an organic concern for students before priming them with a formal definition of assessment in the classroom. Out of the 13 students interviewed, 10 of them referred to assessments in some way as their main concern on the first day of any course. For some, it was as explicit as needing to know the "grading distributions" or "most importantly, when the exam is," while for others it manifested more subtly as "when the homework is due," or "the structure of the course," for example, "like how many exams." The second most cited concern for students was gauging instructional style and getting an idea of how to manage time over the course of the Quarter. Thus, assessment was clearly at the forefront of students' classroom experience even prior to drawing students' attention to the assessment focus of this particular study.

The anticipation of course assessments was most obvious when students were asked to reflect on how assessments or the thought of assessments made them feel. From most cited emotion to least, students reported the following in response to assessments: anxiety (7), stress (5), fear (2), burnt out (1), shame (1), intimidated (1), anger/betrayal/resentment (1), good (1), and confident (1). Overwhelmingly, students' affective response to assessments or the thought of assessments were negative, with reactions as strong as "betrayal" and "fear." In a rare contrast, one student said they felt "good," especially when an "assessment reflected what you know."

In order to understand what assessments are currently being employed at UCLA—following the formal definition of assessment provided to both students and professors—all participants were asked to report what types of assessments have been most *and* least frequently used to assess learning in their classrooms. The most frequently reported assessment used in classrooms was the "traditional" midterm exam followed by final exams: typically, in multiple choice format (See Figure 1.1). *iClicker* questions (or their equivalent i.e., Poll Everywhere)—synchronous questions asked out to the class during lecture where students answer individually to gauge understanding—was the third most cited response. This was followed by quizzes and papers (including research papers, non-traditional forms like Op-Eds and reflections, and minute/exit papers). In contrast, least frequent assessments as cited by both professors and students included papers and essays, quizzes, portfolios of work, and cold calling—the practice of randomly choosing students to answer questions during class (see Figure 1.2).

Given this general lay of most and least common assessment practice, it is important to understand (particularly relative to the more negative perceptions of classroom assessment practice) that much of this perception stems from the more prominent *summative* assessments carried out in the classroom. Thus, while a definition and example of assessments in the

classroom was provided for all participants (so as to be on the same operational 'page'), it appears that majority of participants interpreted assessments as those summative in nature (such as exams). This will be discussed further in each of the finding subsections.

**Figure 1.1**
*Most Frequent Reported Means of Classroom Assessment (n=22)*

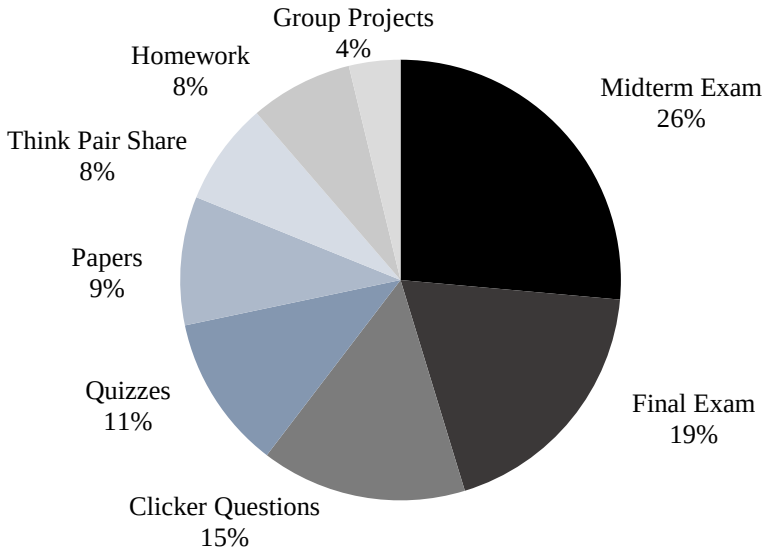
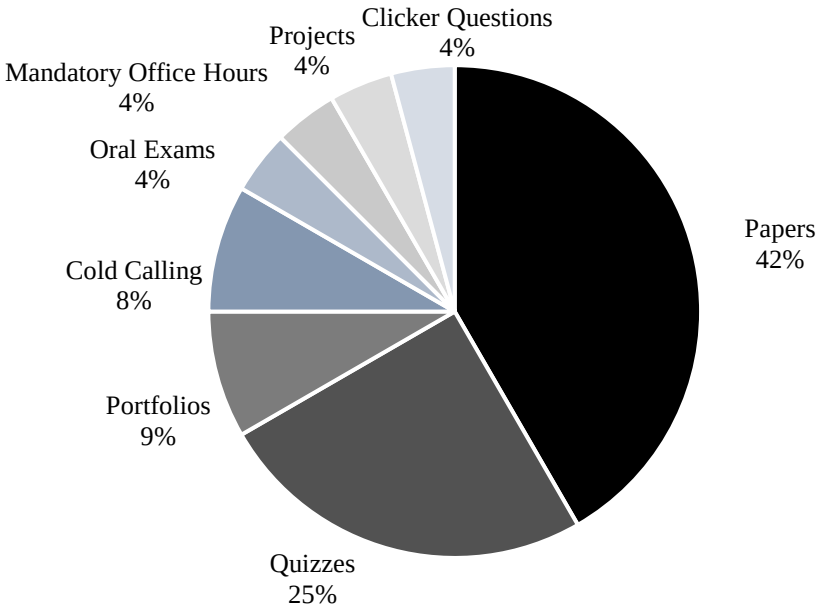



**Figure 1.2**
*Least Frequent Reported Means of Classroom Assessment (n=22)*

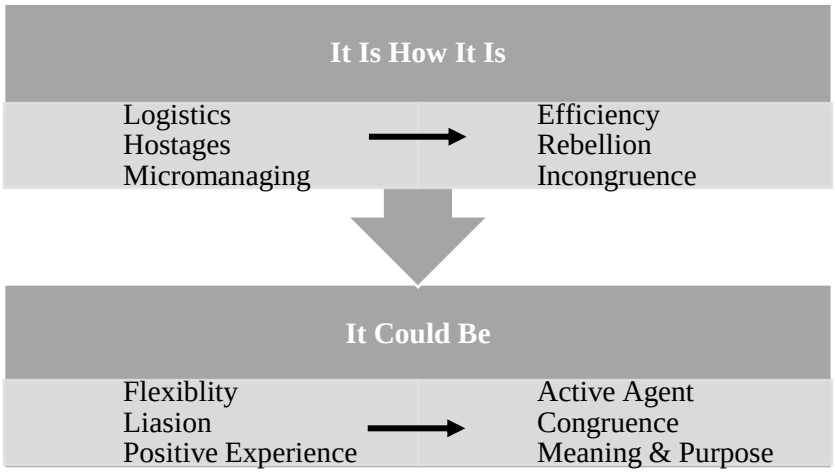


### *Note on Findings Format*

Given the critical lens through which this study was conceptualized and carried out, disproportionate power between instructors and students regarding assessment was certainly anticipated. Thus, an initial conception of disseminating findings was a staggered approach in which student experiences were shared followed by those of instructors or vice versa. However, early on in the interviewing process it became clear that students and instructors share many of the same perceptions of assessment in the classroom (albeit in their own ways). Thus, rather than presenting participant experiences as instructor *versus* students—which would arguably perpetuate the divide between student and instructor voice in classroom assessment—themes will be discussed in terms of how they are experienced and expressed as instructor *and* students. This format is meant to not only more accurately describe the ways in which instructors and students experience assessment—simultaneously rather than disparately—but also to project and model the aim of this overall dissertation: assessment as a dynamic, two-way conversation between students and instructors.

Given the structure of presenting findings as a shared experience between instructors and students, this study's findings can be divided into two main threads (overarching *categories* as defined by Saldaña, 2013): "It Is How It Is" and how "It Could Be." Within each thread, current practices and perceptions are followed by corresponding consequences and implications. Three main themes emerged and are discussed throughout each thread category and its corresponding observations and experiences, including: 1) issues of implicit and explicit power, 2) the foundational need for trust, and 3) motivation in flux. Figure 3 displays a visual indexing of these findings. Interview data are later corroborated with analysis of classroom observations and syllabi analysis.

**Figure 1.3**
*Index of Qualitative Findings: Experiences and Corresponding Consequences*




### *It Is How It Is*

The first thread of these findings relating student and instructor experiences of assessment in STEM courses at UCLA encompass the *current* practices and perceptions of assessment practice; in other words, "it is how it is." Here, participants describe ways they are actively experiencing assessment. The main experiences discussed in this thread include: logistics, hostages, and micromanaging, followed by the corresponding perceptions and consequences of these experiences, namely: efficiency, long-term effects, and incongruence.

***Logistics.*** Logistics in this context refer to the multitude of structural management experiences that participants noted relative to classroom assessment. The most obvious logistical experience was that of classroom size. In one example, a professor described how she did her best to assign writing assignments in order to vary the assessments in her course, but with 150 students and only one TA, felt she was "pushing the boundaries in terms of grading." Similarly, relative to participation evaluation and the use of *iClicker* poll questions, a student noted "there probably wasn't a better way of assessing [participation] during lecture of 150-200+ students." Thus, not only was course size a logistical concern raised by both students and professors, but so

too, was resource distribution with TA's only being assigned one per every 150 students in some departments. When asked how they have considered bringing students into the assessment conversation, a professor noted:

> "with 225 students, I don't really have many alternatives on how to assess them. So…if they give me some really good ideas most of them are not implementable in my context. So, if you ask them for their ideas and then you're not able to respond to that it seems like why are you bothering to ask?"

In this way, participants acknowledged the constraints of classroom size when assessing in ways other than traditional multiple-choice exams and polling software, in addition to how this poses a challenge in thinking of novel ways of assessing: including bringing student voice into assessment conception.

Classroom size not only referred to numerical number of enrolled students, but also the literal and physical size of each classroom. As noted by an introductory professor in Psychology:

> "the environment that the students take the exam in…we have a huge classroom—giant. It feels so anonymous and you just feel like…a tiny fish in there and the idea of having a difficult experience on an exam impacting students in that way makes me wish I could even just put them in smaller rooms…"

A student cited that during exams he felt "we were like sardines" packed into the classroom—a very literal reflection of what this professor perceives as students experience being "tiny fish" in this "sea of students." These logistics of classroom size clearly affect classroom experience as a whole, but also specifically that of assessment experience as noted in the latter testimonies.

When asked about first day concerns a student mentioned grading distribution relative to her field where "only" 25% of students received A's and 35% received B's: "even if I do well…there's a chance I might not get it [an 'A'] just because like, of the competitive nature of the major." In this example, the student inherently defines a competitive major as one that is exclusive in its grading distribution. The fact that "*only* 25%" of students receive an 'A'—

regardless of the fact a student has high scores but simply does not make it into the top quartile—is not only seen as a potentially hopeless venture (hence "even if") but also seems to suggest to students whether a course is designed to be collaborative versus competitive in nature. Thus, before students even begin learning course content, the mere logistics of a grading distribution can suggest to students 1) whether they are capable of succeeding, and 2) how they might perceive their peers in the context of that class (as collaborators versus competitors).

A final logistical concern was that of objectivity in assessment. When discussing the challenges of providing alternative assessments in one course, a professor cited: "It can feel like comparing apples to oranges when you're providing different opportunities…[there is] lingering discomfort that I felt the entire time…because it was so unfamiliar." Several professors hinted at the need for objectivity within their assessments. This came as no surprise, as the perceived objectivity of STEM disciplines appeared to seep into ('what should be') the perceived objectivity of certain assessment methods. In this example, traditional assessment practice appears to satisfy this pursuit of objectivity, while alternative assessment methods (like providing students the option to choose test format) seems to trouble this notion. This then led to feelings of "discomfort" that "lingered" as the professor described this instance of attempting to provide flexibility and student choice in assessment practice. While it is unclear if this was explicitly the case, this sense of "discomfort" may also have derived from a shift in assessment power, where student voice was now being included in assessment practice.

***Hostages.*** While grading distribution was one logistical concern shared by professors and students, it also doubled as a mechanism through which participants viewed and/or presented themselves as hostages. When asked about conversations in their department that address assessment, one professor noted: "there's an unspoken expectation [of]…variability in the

grades…my belief is that if I'm doing my job…everybody could and should earn a…high score in the class. But I feel—and no one has ever said this to me—it was—this is implicit." While this professor doesn't name the grading distribution as such, she hints at the implicit expectation of her department that her larger courses will have "variability." This professor's struggle to articulate her feelings of what her department implies in addition to her continuous efforts not to exert blame on her department ("*I* feel," "no one has ever *said* this to me," "*I* feel like"), exemplify an inner struggle between an instructor's own intuition and one's department, especially given her positionality as a newer, non-tenured faculty member.

In a more explicit example of faculty expectations, one professor recounted something she had overheard in another (non-STEM) department:

> "The Chair…said, 'your teaching scores are way too high and that means students value you…you probably are putting way too much time and effort into teaching, and I'm going to guess that that means you're putting less time and effort…into your research. That's going to get you into trouble.'"

While this example is not exclusive to classroom assessment nor STEM, it is indicative of the larger culture in which instructors must develop both the assessment and overall classroom experience for students. As in the previous example, this was describing a new, non-tenured faculty for whom, perhaps, the expectations were thought to have to be made explicit. In this instance, the professor was reminded of the importance of one's research identity/work at an 'R1' research institution *over* their instructor identity/work. The fact that the professor had glowing student evaluations suggested to his Chair that he must have been sacrificing his research (primary expectation) for the sake of teaching and the student experience (secondary expectation). In short, this professor was hostage to a culture where student experience is expected to be average or low.

Students were no less hostages to the system and appeared to specifically recount such experiences relative to more 'traditional' assessment forms such as exams. When asked about assessment experiences that stood out to students, one student, a Junior with a double-major in both a STEM and other field, responded:

> "There's this whole midterm season, and it's like people are dying…[we're] not sleeping or everyone's sick. It's just like, this terror…literally all you have to say is midterm season and everyone understands and are like "yup," because you know how stressful [it] is…this is like life and death situation…like, if I do bad on one then that's going to determine my whole grade for that class."

The initial question was asked neutrally in order to probe any type of experiences that students most remembered from their assessment history at UCLA, both positive and/or negative. Generally, however, as in the case of this student, most student participants cited a negative experience with assessment as the ones that stood out to them. This student describes very explicitly the trials of "midterm season" as it is known: "not sleeping…everyone sick," "stressful," and "life and death situation." In this particular example there is a very clear negative, dreadful connotation to "midterm season," that appears to be common knowledge among students.

Relative to exams, one professor noted they "see exams as kind of a necessary evil"—not something they believe is the *best* way to gauge students' learning, but one they must engage in, often as a result of logistics such as class size. When asked about her most enjoyable assessment experience, a professor relayed the formative or "ah ha moments" that are enjoyable and motivating as an instructor, rather than exams. She closed with: "Nobody ever comes back and was like 'Your test was the best test,' [*laughter*] um, and so, I wish there were a way to…capture that…in the classroom." This collective surrender on part of participants to the "necessary evil," of assessments (particularly those summative in nature), further suggests this

hostage existence that both students and professors appear to experience. As this professor comedically notes, students don't, in her experience, come back to her as an instructor and laud the tests she has administered in the past. In this, she recognizes that testing is not an experience that students recall fondly.

These summative experiences often disclosed the passivity with which students approached their experience in the classroom. For example, when asked to what extent a student perceived having power in the classroom, one responded that relative to exams: "I don't have power…I've kind of accepted that…so I don't have any power in making those decisions." This student has come to terms with her lack of power in the realm of (summative) assessments and framed these as something that she just "has to do." Interestingly, however, she goes on to say that she doesn't think this experience is limited to herself as a student: "I don't even think sometimes professors have the power in making those decisions… [it's] department things…so I don't blame [them]." In her rationalization of lack of power in classroom assessment, the student empathizes that perhaps even professors themselves don't have power in this regard, and thus, defers from putting "blame" on them. In this way, the student purports that for both students and professors, this simply 'is what it is,' relative to power and classroom assessment: highlighting this helpless hostage positionality.

A nuanced way in which the hostage experience revealed itself was through a combination of professors citing esoteric authority over assessments, while students simultaneously deferred to such authority. In one example where a professor attempted to provide the opportunity for students to create potential exam questions, he claimed: "I don't think it's a good idea" as students often wrote lower-level cognitive questions. Similarly, another professor talked about the experience of providing flexibility in project topic to students citing

"mixed feelings": "giving…agency for some…means they're just going to do something that's…easy…it's like relying on a student to make a choice that will benefit them in the way that you wanted." In these examples, professors cite a lack of students' metacognitive ability and foresight in being involved in the assessment process. This follows, given that students are not often involved in assessment development, and thus, would not necessarily have the skills to engage the way that instructors, with their varied experience, do. In each example, the professors hesitate to share power with students as they do not have confidence students will engage in the way professors intended. This esoteric stance from which many professors approach assessment practice perhaps explains why shared power in the assessment realm is seldomly realized.

The esoteric stance appeared to translate to the extent to which students perceived *they* had the ability to be involved in assessment practice, too. In an explicit instance, when I asked a student if they would want to be engaged in assessment dialogue, he said: "You will have read studies [and] know more about what will actually help…I'm probably not the best person to…I wouldn't—wouldn't want to do that." In an unexpected turn during the interview, the student participant addressed *me* (the researcher) as an expert in education or education research, such that *I* would be better suited to make such decisions and be involved in such change, rather than himself. His fragmented response demonstrates the hesitation with which he was considering the opportunity. This response was the most explicit instance in which a student deferred to power (in this case, myself as a researcher). This habituated deference almost across the board with students, pointed to a deeply ingrained perception that engagement in assessment was either not in the students' realm of capacity or responsibility. To sum the participant experience in the words of one professor, "we are really hostage to the system. That's what I look at this as..."

***Micromanaging.*** While professors and students describe many of their assessment experiences as that of being a hostage, their experiences were certainly more than just passive. Micromanagement emerged as one way both professors and students appeared to cope with assessment in the classroom. In one example, a professor described how asking students to attend class "just on their own merit" when lecture recordings would be made available led to a decrease in student attendance. A common practice in large STEM classrooms is the supplement of "BruinCast," a recording of each lecture which appears on UCLA's online learning management system shortly after each class. This professor notes in the courses where this recording was available, students would opt to watch the lecture online instead of attend class in person. The professor began implementing *iClicker* questions during class as extra credit in order to "incentivize" attendance, citing that it not only helped with attendance, but also with "overall performance" in the course. This type of micromanaging by professors was not uncommon. In another example, however, this professor notes a potential issue with such a scheme:

> "I don't know what's going on in the lives of all 275 of my students…that became really clear to me during …the fires…students couldn't come to class because they were in areas that were like, literally on fire…But that made me really aware…would I want to drive…an hour and a half in traffic just to listen to some lady talk about something that I already understood in the book? Probably not… so I don't want to punish them"

This professor appears to describe an inner struggle that comes down to an issue of equity. As a result of fires one Quarter that made it near impossible for students to attend class, the professor realized how far some students were commuting (and thereby the time and other resources students were employing to get to campus). Putting herself in their shoes she noted that if she could get the same information from the textbooks, she might not likely attend lecture either. Thus, while she has data to suggest that some students end up doing better as a result of incentivized, in-person attendance, she doesn't want to inherently "punish" those who cannot or

choose not to. In this example, the professor sheds some light as to the reasons for, as well as potential challenges of, micromanaging student attendance. There appears to be a to-and-fro of wanting to reward student behavior essentially for their own good (evidence of better performance) and wanting to acknowledge and be equitable toward the student experience.

Just as professor experiences were mixed in this regard, so too, were student experiences. One student noted as a response to some of these attendance policies that "life happens and…that's part of being an adult…it's tough…but in college it just feels like I'm [treated like] a kid." She went on to say that she wishes the system were based on "mutual agreement" where students are held to standards they create such that "it would teach students" about "responsibility" and "humility,"—essentially, "how to be people in the real world…" This, however, was contrasted with another student who noted "great value" in micromanaging practices "because if we weren't assessed and we were just going to lecture every day, I would be pulling out my phone and I would be sitting there and not really caring, right?"

While in the former experience, the student seeks autonomy and being treated like an adult, the latter experience speaks to a controlled motivation that relies on incentives in order to stay engaged. Assessment purpose, to the former student, should be geared towards preparing students for life beyond college, while for the latter, is geared towards the micromanagement of student behavior to encourage learning. This differential meaning making of assessment purpose will be expanded on in a later section. To sum—quite frankly—in the words of one professor: "it's extremely easy to manipulate student behavior with point schemes…so I'm always thinking about clever ways to do that."

Finally, micromanaging in assessment practice came up most often in regard to cheating and academic dishonesty. While several professors cited cheating as a concern in their

development of assessments (particularly those summative in nature), one professor in particular, discussed it at length. When asked about what he might normally be discussing with colleagues regarding assessment, among other things he mentioned conversations "[are] probably going to center around like cheating on assessments":

> "If you don't want people to cheat you have to…control their assessment…[it has] become much more complicated…because of COVID. Assessment has become the primary challenge in my department and many…STEM [courses], because…the primary mechanism of doing difficult-to-cheat-on assessments is gone…And it has clarified for me why…I feel even more strongly that without in-person instruction and the ability to control assessments…it's almost impossible to achieve the same educational outcomes…you might be able to use assessment as a sole justifier for the existence of the physical university"

The sheer amount of thought and emphasis on micromanaging assessments such that it makes it hard to cheat on for students stands out in this professor's experience. The use of language such as "control," "primary challenge," and feeling "even more strongly," show how much of a concern cheating is when it comes to assessments. This concern is exacerbated particularly in STEM fields where assessments are often in the form of summative exams, and more so as a result of COVID-19 and the switch to online instruction. In fact, the professor feels so strongly about cheating such that he suggests preserving "the existence of the physical university," could rest on the "sole justifier" of preventing cheating on assessments. This professor feels that the only way to maintain true academic integrity is for students to take assessments in-class under the proctoring of a professor/teaching assistants. Only under such micromanaged conditions, does the professor believe his course can "achieve the same educational outcome."

Logistical roadblocks, feelings of being hostage to the system, and micromanaging were all experiences that lent themselves toward participant perceptions of incongruence, a strive for efficiency, and eventual rebellion towards existing assessment practices.

***Incongruence.*** Incongruence took a variety of forms in the experiences of professors and students when it came to classroom assessment—from incongruent perceptions of what students felt they deserved as a letter grade versus what they actually received, to professors' incongruent perceptions of providing student autonomy and responses received. In an example of a student's experience with exams, one student lamented about some of his poor grades: "oftentimes [it's] my own fault. I'll study things deeply and try to spend time to understand them and then run out of time…versus had I just like, memorized the right answer…this bugs me 'cause I feel like, I'm just like, playing the game and I don't like that."

This student suggests an incongruence between what they believe to be the purpose of studying for an exam and what is potentially the more strategic route. This student brings to life an inner dialogue about ownership of one's own performance (it is his "own fault"), in addition to a conflict of interest. While the student realizes that switching his strategy to memorizing material might benefit his grades, he sees it as incongruent to his current aim. That is, memorizing material is "playing the game," and can lead success in terms of grades, but this contrasts the students' own preference for studying. To him, assessments revealed an incongruence of purpose: whether to "study things deeply"—which leads to poorer outcomes—or simply memorize and perform well.

Incongruence was also cited by a professor relative to their department's views on the purpose of assessment. When asked what he wished his department would know about his experience or improve on, he said:

> "more discussion and agreement about the role of assessment in learning…some people—and they're not ashamed of it…have a different point of view… I've had people say "Oh I love my exam because the average correct was only 50% and that's really great…I had a nice Bell distribution,"…and I'm thinking, I

> can't imagine giving an exam and being happy if only 50% of the students could get a question!...I don't think "Oh great, there's a nice differentiator."

Here, the professor cites the incongruence of assessment purpose within his department—with some professors seeing lower distributions as a positive outcome of the assessment, while he and others see this as a detriment to their teaching. From his perspective, some professors do not see the "role of assessment in learning," but rather as a mechanism of differentiation (and are "not ashamed of it").

Similarly, students noted a mismatch between instances of being asked to bring their voice to classroom matters and its *actual* usage/effect. When asked about instances in which students have felt power in the classroom, one student noted:

> "the evaluation that we're asked to do at the end [of the quarter]…I'd like, argue that that in a sense isn't even power. [It gives] professors something to go off for next quarter, but at the end of the day it's not doing anything for us…Our learning experience has already happened"

In this way, the student notes an incongruence between being asked for her voice on classroom matters and it having a noticeable outcome of her own classroom experience. This incongruence of contiguity—while seemingly an instance in which student autonomy is encouraged—doesn't appear to make much of a difference to students because of its retroactive nature.

Another instance of incongruence which students claimed was a compelling reason for which their voice should be incorporated into classroom assessment practice, was the difference in student and professor classroom experiences. She cites "it's a really good idea for students to be more part of it" as their experience differs from their professors' student experience: "things have changed [since then]—the way we study, the way we learn [has] changed drastically so [student input] in those assessments I think would make a difference." Here, the student cites incongruence between what previously *was* professor experience as students and what *is* current

student experience: and the need for current practice to align with current student experience. Moreover, there appears a desire on part of students to want to counter what has been the assessment status quo in order to cater to various student learning styles.

Relative to incorporating student voice into assessment practice, one professor noted a perception of students:

> "which is kind of putting all the credit into something easy—I don't think that's true. Actually, I think when you give students the opportunity to [say] how they want to show their knowledge, they do a really good job…My main concern with that would be like, how do you take all the differing opinions of all the people in your class, right, and come to some sort of consensus about the grading scheme you want use?"

In this excerpt, the professor cites two different types of incongruence. Firstly, an incongruent perception on part of some professors where asking students to engage in the assessment development process would mean students looking for the "easy" route. In her experience, the professor has found the opposite such that students provide a variety of assessment categories that best represent their knowledge (not just that which is easiest). Secondly, the professor cites an incongruence between the *concept* of student voice in assessment practice and its *implementation*. While the concept can lead to students demonstrating their knowledge in new ways, how does a professor consolidate what will likely be differing perspectives on what classroom assessment should look like? Thus, while this professor sees value in taking student voice into consideration in assessment, she struggles with what this would look like in practice.

A final incongruence between professors' intention when being flexible with assessments, and subsequent student reaction was described as such:

> " I think [students] are very grade oriented—so…I think there is always this underlying like, 'I will write whatever it takes to get the 'A'…[they're] kind of stuck on that idea like, well, 'what is the right answer?' and there's not a right answer. There could be many right answers, but they don't like that."

Here this professor expresses an incongruence between their intention when creating a flexible assessment: allowing students a space where there is no *one* right answer in order to allow student voice to shine, and students fixated on what it is that will earn them the 'A.' While this may seem like a harsh evaluation of students, I caution the interpretation of students in a negative light. With an assessment history of feeling hostage to the system, it follows that students lack the expert knowledge in the assessment realm. Having likely always had professors or teachers provide judgement on their work, students may not have had the chance to develop an internal gauge of their own work leading to this dependency on the professor as to what the "right answer" is, which here, appears to foil the intention behind the desire to engage students in flexible assessment practices.

In all, classroom assessment practice in its current state appears to lend itself to incongruent perceptions of assessment purpose, student voice versus actual practice, and more.

***Efficiency.*** As a result of the logistical concerns and hostage-like experiences, participants tended to justify their actions in the frame of efficiency. In almost the same words, several different professors describe the ways their students were motivated:

> "students are motivated to get an A in the class…"
>
> "I think students are highly motivated by those assessments that contribute to their grades."
>
> "there's like a lot of motivation to like not get a bad grade…"

These strikingly similar comments, that too, from professors in varied STEM departments, prefaced much discussion of the ways in which assessments motivated students.

One professor noted:

> "…you only have bandwidth for so much in your life…I can see where you have to make choices in how much effort you dedicate to any part of the class. So, if it's not gonna count toward a grade, and grades are so important in this culture of school… 'let me go focus on something that does.'"

In an attempt to acknowledge and rationalize student behavior, this professor describes the various ‘pulls’ in any given students’ life which might explain their explicit motivation towards assessments that contribute to their grades. In doing so, the professor also points to the systemic “culture” of grades in school rather than individualizing the drive and placing blame on students. She went on to describe, however, that “it’s a little frustrating that…they don't want to learn for the sake of learning. That if there’s not an assessment attached to it, they're not as motivated to participate in and engage.” Thus, while the professor understands student aim of efficiently allocating efforts in a course, she cites the frustration and challenge from the instructor’s point of view in attempting to engage students throughout the academic Quarter.

One particular student interview sums the efficiency with which students tackle their assessment experience, and how that can impact how students feel about learning. When asked about her most enjoyable assessment experience, this student claimed:

> “we are kind of built to test since, you know, like, high school so…[like] I have a really good memory…[to] cram and then on the test, for my assessment to be kind of predictable. It could just be like a multiple-choice type of assessment…that would be in some ways most enjoyable…like I know I could perform well on it. But I would say…enjoyable...doesn't necessarily mean the most enriching.”

When asked at the end of the interview if there was anything we hadn’t discussed that she would like to share relative to her assessment experience, she reflected:

> “the fact that my natural response was that enjoyable means kind of predictable…does kind of make me feel…[a] bit like disheartened that works…But since…my whole time in education is…I've been taught to just prepare…which doesn't always lead to retention of the information but instead, learning how to take tests really well”

This unprompted reflection and reckoning with the way she answered her most enjoyable assessment experience was truly insightful. The student recognizes the dichotomy that predictability of assessments and those where memorizing maximized her performance is

enjoyable but not necessarily as "enriching" as it is "disheartening." This emphasis on performance (since at least high school in her case) has led her to develop skills that are efficient in helping her do well on "multiple-choice type of assessments." However, later she notes these same skills are likely not helping her retain information but have just lent themselves in making her a good test-taker. The fact that the student came back to address what she felt was a discrepancy in an answer to a question from near the beginning of the interview (that too, without probing), demonstrates an internal conflict of one's feelings reflecting something (enjoyable, as a result of being efficient) that her personal values might not (learning for the long run). This reflection was reminiscent of students' hostage experience of playing the assessment "game," and often, being disappointed they were doing so. Thus, while students are clearly efficient in developing the skills that will help them perform (communicated through assessments as being important), they are not necessarily content with doing just that.

To sum these experiences in the words of one professor:

> "I wish [there was] engagement with the material and learning as opposed to 'I need to get an A because I'm going to Med school'…And I totally understand it…so like I don't know if it would be a matter of like, abolishing grades…But I don't really see an automatic solution here…especially at UCLA. But I think I would love to see like, assessments based on engagement and understanding of concepts for…the joy of it, as opposed to avoiding a negative outcome of not getting a good grade."

While efficiency was a clear strategy used by students to tackle the various logistical challenges, micromanaging, and incongruence in their assessment experience, so too, was this the case for professors. When asked about what might be used to encourage faculty to incorporate student voice into assessment practice or engage in flexible assessment, many professors voiced something to the effect of it being part of their department values. For example, one professor suggested: "something that is part of your merit and promotion package," while another noted:

"[convincing faculty] that their [student] evaluations would increase… [as] evaluations hold a powerful position in faculty tenure so [they might]…be responsive to."

In the same way students exhibit efficiency in their efforts and attention in order to perform well and achieve certain grades, professors too, exhibit efficiency in their desire and attempts to alter their assessment methods relative to its effect on their career trajectory.

A professor, in her second-year as a faculty member on the tenure-track, expanded on this phenomenon when asked about the types of conversations she has with colleagues about assessment citing:

> "being at UCLA which is a school that your time as faculty is primarily meant to be on research…[we are] riding out this fine line between doing enough as a teacher and having enough time to do your research [which] can often result in…cutting corners…my colleagues know a lot of the best practices when it comes to assessment but…[there are] barriers to implementation whether it's time or class size."

In the same way that the student in her reflection made a distinction between what works for her performance but may not be "enriching," this professor makes a similar differentiation between what they know to be "best practices" and their "actual" practices. As a result of balancing various responsibilities and being at an institution where the primary focus is research (another hostage example), the professor admits that faculty may "cut corners" to meet demands in the same efficient manner students might "memorize" or "cram" to meet the demands of their assessments. In this example, the professor describes her vision for alternative assessments and how, as a result of "barriers to implementation" (logistics), she is forced to narrow her scope.

Exploring both student and faculty experiences of employing efficiency side-by-side (rather than in isolated silos), begs an empathetic, rather than scrutinizing perspective to participants' actions in the realm of classroom assessment. From this lens of shared experience, it becomes clear that desire to learn for the sake of learning and to queer assessment practice does

exist but is challenged by a variety of factors that appear to favor controlled motivation in the form of efficiency and "cutting corners."

***Long Term Effects.*** From the combination of logistical concerns, hostage experiences, and micromanaging, both professor and students touched on the long-term effects of their classroom assessment experiences at UCLA. As I posed the question: "Can you describe a few of those [assessment] moments that stick out to you the most?," a student whispered, "so traumatizing." And while this comment was in jest, the extent to which classroom assessment had affected her and other participants should be considered from serious vantage point.

In many cases, students cited fear when they thought of classroom assessment. For example: "these tests can basically dictate…my future," while another quoted: "it's a lot of fear by not performing well there will be negative outcomes later on in life." Classroom assessment was not just about their present performance in a particular classroom, but rather, the permanent implications that performance could have on the student's future opportunities. More specifically, for one student: "I dropped out three times so it's…the fear of failure" where assessments led to "feelings of inadequacy" or not wanting to "make a mistake." For this student, having dropped out of college several times and being of non-traditional college age made assessments seem like permanent judgements on his "adequacy" and ability to remain in higher education. This "fear of failure," led to the perception that "mistakes" relative to assessment were something to be avoided rather than opportunities for learning and improvement.

This hostage experience lent itself to fear and trauma for many students, while for others, it served to exacerbate their hopelessness and, in many cases, decrease student motivation. Relative to her negative experience with midterms, one student cited that they are so heavily weighted: "if I mess up on one test…it's like, demoralizing….there's a very slim chance that

[I'm] going to be able to recover." The overemphasis on midterms for grading is accompanied by an overemphasized perception of students' own ability to succeed. This controlled motivation to perform well on exams turns into a total lack of motivation in the long run, if the first does not align with students' definition of performing well.

One professor corroborated this exam experience, citing: "[I've] encountered a lot of students who feel pretty demoralized…the mere fact…[they] felt like they failed [can affect] their self-concept as a scientist." Coincidentally, both this student and professor use the term "demoralize" to describe the effect of not doing well on an initial exam in class. Here, professors realize the hopelessness and crippling effect of doing poorly on a preliminary assessment and recognize its long-term effects. In this case, the professor sees how underperforming can negatively impact students' "self-concept as a scientist"—a consequence that may be heightened for underrepresented students in STEM.

Demoralizing experiences with assessment led to differing outcomes for students. In one example of a student in her Junior year—one who had transferred out from her original STEM major due to poor performance—said: "I shrug it off because my GPA is ruined…I could fail and like, I wouldn't feel any remorse because…I've already messed it up." In her experience, the student feels that her prior performance has "ruined" her GPA such that there is no return; in her words "I've already messed it up." This hopelessness led her to adopt a blasé attitude such that grades were now meaningless to her.

In contrast, this hopelessness in some cases led to a complete dismissal of the status quo. One student cited her initial years of frustration with classroom assessment and how "eventually that anxiousness and anger…just mixed together, and I had the sense of not really caring anymore…[It was] almost like, a sense of rebellion." Much like the previous student citing no

care for failing, this student cites how the thought of an assessment no longer provokes the overwhelming feelings of anxiety and anger as they did in her early years. In fact, for her it led to a "rebellion" against the traditional ways of being assessed in her classes.

She went on to describe this change in her perception—as "part of this broader philosophy change"—where she began to take an interest in learning about learning and developed "[her] own standards." Her new mantra was: "I'm going to apply *my* standards and not *your* standards." Part of this student's rebellion led her to reject her professors' existing assessment standards, and instead, seek to understand the research on learning and apply those findings to herself. This interest stemmed from her experience as a "Learning Assistant" (LA). This program, sponsored by the Center for Educational Innovation & Learning in the Sciences (UCLA, 2021), recruits undergraduate students who have taken introductory STEM courses and provides pedagogical training for them to facilitate learning for other undergraduates currently taking those courses. For this student, the LA experience accompanied her change in long-term perceptions of assessments and the purpose of learning.

While for some students the long-term effect of their assessment experience led to rebellion and adoption of personal standards for learning and performance, for others it had more dire consequences. One student described her early experiences at UCLA with respect to her former aspirations of pursuing Pre-Med in her first year:

> "[I had] 'weeder courses,'…they're literally trying to discourage students…[it's] so messed up…they don't need to make those classes so unnecessarily hard…that shouldn't be the reason why [I dropped out]…The reason should be 'I'm not passionate about this'…it affect[ed] me so negatively…I just gave up on it."

This was an illustrative example of the idea that classroom assessment can sometimes affect students' "self-concept as a scientist." In this case, the impact was so negative that the student "dropped out" of the Pre-Med track. The term "weeder classes," is typically used to describe

courses in which only a certain number of students are expected to succeed, otherwise described as “sink or swim” classes (Koebler, 2012). By and large, such courses inevitably result in disproportionate impact to marginalized populations in STEM, such as in this example, the intersection of women and people of color. While in her conception, college should be a place to “figure out” one’s passions, the difficulty of the course—as communicated by the classroom assessments (in many cases, a ‘standard curve’)—can suggest to students that they do not belong or do not have the ability to succeed. This hopelessness and “unnecessary” difficulty led the student to give up, which she cites as a fairly common occurrence with students at UCLA.

While much of the long-term effects of hopelessness and trauma were communicated by students, professors cited equal impact from their classroom assessment experience. In one example, a professor noted an interaction with a student after one of her exams where: “there was a student crying and she made comment about self-harm. And I felt like, ‘If my exam can do that to someone, then I'm doing something wrong.’” While the exam was clearly traumatic enough to cause thoughts of self-harm for the student, this assessment experience also traumatized the professor as evidenced by her emotional narration and given that this was the one instance that stuck out to her in her reflection of memorable assessment experiences.

In this way, classroom assessment experience, in its current form—particularly the part that is summative in nature, is clearly more than just an isolated and temporary experience for both students and professors. In many cases, it led to a variety of long-term effects for both students and professors ranging from trauma to helplessness, and rebellion. For all parties, classroom assessment represented a figurative and/or literal extension of the ability to succeed as well as opportunities for the future.

### *It Is How It Is: Summary*

This first part of these findings of student and instructor experiences of assessment in STEM courses at UCLA documented *current* practice and perceptions. In describing "it is how it is," participants mentioned phenomena relative to logistics of the classroom, feelings of being hostage to the system, and micromanaging of behavior. These phenomena led to perceptions of incongruence (whether of expectations of assessment purpose or in considering the implementation of student voice in assessment practice), a strive for efficiency (whether through students exerting effort to those assessments contributing to their grade, or professors exerting effort in those practices that contribute to their teaching evaluation/tenure), and finally, long-term effects of assessment practice (whether leading to general hopelessness or complete rejection of current assessment standards).

In the words of one student:

> "even though grades might not be the only thing that matter…we have really been taught over time that in order to succeed you need to perform…motivation plays a part in wanting to study and pay attention…because it's a lot of fear by not performing well, there will be negative outcomes later in life."

In this example, the student expresses an incongruence with the belief that grades may not be "the only thing that matter," as a result of the academic socialization (read, hostage experience) over time that grades are necessary to "succeed." This led to a strive for efficiency by exerting controlled motivation to "grasp information and pay attention," in order to meet this expectation of "performing well" (read, micromanaging). All of this, which leads to "fear" of potential "negative outcomes" in life, as a result of the way assessment practice informs future endeavors (read, long-term effects).

While the findings heretofore have addressed current assessment experiences, this leads to the second half of findings where instructors and students express the future of assessment practice in terms of what "it could be."

### *It Could Be*

The latter thread of findings relating student and instructor experiences of assessment in STEM courses at UCLA touch on the potential future of assessment and perceptions of assessment practice. In "It Could Be," participants describe the possibilities for assessment practice in addition to reflecting on how such practice could affect their perception and motivation. The primary topics discussed in this thread include: flexibility, the liaison, and positive experiences, which lead to the following perceptions and consequences, namely: being an active agent, congruence, and overall meaning/purpose.

***Flexibility.*** While many participants spoke of assessment in strict terms, they also discussed the ways in which their assessment experience has allowed certain flexibility. While collection of mid-Quarter classroom feedback was the most popular form of flexibility cited by students and professors (not restricted to assessment practice itself), a similar technique of soliciting feedback *specific* to classroom assessment was also cited.

One student recalled a time where the professor asked students which format they would prefer to have their midterm exam. The majority chose multiple-choice and she described the experience as "beneficial": "[it] showed her understanding like 'OK what is the best way you guys are gonna like be able to learn?'…'cause we, we know how we're best assessed…[she took] the time to actually ask…rather than assuming." The student recalls finding this flexibility as "beneficial," in the way that it demonstrated a desire to understand student perspective "rather than assume," and remain entrenched in 'it is what it is.' Interestingly, the student refers to the

thought process behind this as postulating how students would be able to "learn." For this student, the test was a mechanism for learning, rather than simply an opportunity for performance. It is unclear whether this perception stemmed specifically from the ability to provide feedback (e.g., 'If I get to choose, then this is a learning process' vs. 'If the professor chooses, then this is simply a performance task').

In a similar way, one professor described a time during one of her summer courses where she allowed students the *individual* opportunity for exam format choice. In contrast to the previous example, whatever format each individual student chose was the format they received. She said: "[it was] what they felt were their strengths. The students were very responsive to that…[they] valued the opportunity to have some autonomy in the way that they would have their knowledge of the coursework assessed." The professor describes the positive response from students in being able to be assessed relative to their "strengths," and highlighting the potential sense of "autonomy" students take away from such an experience. In her description, the professor also describes the limitations of such an approach. Due to offering each student an option, this made things "harder," for her in terms of the amount of work required to make such an option a reality. She later mentions that because it was Summer, she was able to "experiment" more but that the experience made her question the reliability and validity of her assessments (given how format choice can alter the uniformity of assessing students). To her surprise, students did not approach her with similar concerns.

While offering a choice of assessment format was mentioned sparingly by participants, a more common form of assessment flexibility came in the form of multiple grading schemes. Nine total participants, both students and faculty, mentioned implementing or experiencing flexibility in grading schemes. One student recounted a class where for a total of three exams, the

professor allowed whichever exam a student scored highest on to be weighted the most, followed by the second highest exam score, and so on. He reflected: while "there was flexibility in the grading scheme, it wasn't necessarily a choice that we could make." This is peculiar practice in that while there exists flexibility, there is no explicit choice on students' part which contrasts previous examples of student choice as part of flexible assessment practice. Overall, this type of scheme appeared to still have positive effects: "It was nice because…[it] set you up best for success." While the student cited this type of practice as "an anomaly," he continued on to mention that it "feels good because it felt like the professor was on my side…[and] being forgiving of like, not doing well." Thus, despite not having an active choice, the student still appreciated the opportunity to have his weaker scores "forgiven," and feel as though the professor was on his "side," in a way that was looking out for his overall "success" in the course.

Part of assessment flexibility also came in the form of assessment variety. A student cited that she appreciated when classes use a variety of assessments: "I feel like it's taking into account multiple data, multiple situations." The variety of sources (akin to triangulation in qualitative research) was something students enjoyed. Another student reflected on "think-pair-share" practices in her class: "[I was] pleasantly surprised…at first, I thought…'I don't want to talk to my neighbor'…[I'm] antisocial when I go to class…[but it's] been one of my favorite ways [to learn]." This variation in assessment methods was a "pleasant surprise," for this student in her upper division courses. More specifically, she appreciated the opportunity to exchange perspectives with peers where students get an opportunity to reflect for themselves, and pair up to share their thoughts. Despite considering herself "antisocial," she notes thoroughly enjoying these opportunities and cited these instances as ones that stick with her beyond the classroom.

In the vein of flexibility, one professor discussed a method he used which varied from convention: two-stage exams. In the 'first stage,' students take an exam individually as they normally would. This is followed by the 'second stage' where students are assigned groups and retake the same exam as a collective. He cites:

> "it's pretty non-traditional for exams…especially high stakes…students tend to have a pretty positive reaction to the team stage and…more intense collaboration…[it's a] mix of extrinsic motivation…wanting to do well on the exam and having to navigate…people as a resource to help you accomplish that goal."

The professor touches on the status quo of exams, particular those "high-stakes," in nature relying on individual performance. The summative reality makes a practice like "two-stage exams" appear "non-traditional." The professor not only assesses students' individual performance via the first stage of the exam, but also seeks to assess students' "collaborative" abilities via the second stage of the exam. He notes that this practice is driven by students' extrinsic (read, controlled) motivation to perform well on the exam, and that given this setup their desire can only be achieved by indulging in the collaborative process. In this way, success is no longer simply an individual phenomenon, but a collective one.

A final example of flexibility that participants suggested assessment practice "could be," came in the form of assessment that had personal relevance. A student describes: "I've had this happen very rarely, but maybe in essay form, applying the concepts to what I'm actually seeing in the world around me." She recalls this only happening in Spring Quarter 2020: "with COVID and the protests. I think…professors were getting more creative and also allowing us to reflect more on our personal engagement." This student refers to the global pandemic (making online instruction mandatory at UCLA), in conjunction with the protests seen nationwide for racial justice. It was only during these times in which higher education was forced to adapt to non-normal circumstances, that the student had ever had professors become "more creative" with

their assessments and allow students to "reflect" on STEM course content in a way that helped her apply them to what she was "seeing in the world around" her.

Whether with the physical context or personal reflection of assessments, students cite the importance of assessments being relevant. This points to ways in which assessment practice can acknowledge and privilege students' esoteric knowledge about their own lives and the way it relates to classroom content, which is in contrast to the hostage experiences in which students were held to esoteric knowledge of their professor. Such personal 'expertise,' might be especially advantageous for minoritized students in STEM who may not see themselves typically reflected in STEM and STEM content. Overall, flexibility—whether in the form of collecting classroom feedback, preference and varying of assessment format, multiple grading schemes, or assessments which seek to have personal relevance for students—was cited as a way that classroom assessment practice can strive toward.

***Liaison.*** Up until this point, experiences of students and professors relative to assessment practice has been discussed. There is one party, however, that has not been addressed or spoken to as part of this study which both students and professors referred to quite often. That is, the metaphorical middleperson/liaison between them: Teaching Assistants (TAs).

Without prompting, students often discussed their relationships with TAs as impacting their perceptions of power and motivation in the classroom. One student mentioned that while sometimes she has "really like, old professors," a "young, good TA can just change [her] motivation." Similarly, when asked whether and to what extent she felt power in the classroom, another student mentioned "if you have a relationship with…your TA…you feel empowered" and it "feels like you have someone on your side."

Professors also mentioned their TAs, but typically when discussing the logistics of assessment practice. For example, one professor talked about how his TA quite literally became the mouthpiece of his students during class. In order to field synchronous questions from students, the professor began using an anonymous web application in which students in his large introductory course could submit their questions or confusions during class: "The TA monitors the Google sheet and…has a microphone which means I can hear them, and the TA raises their hand and asks the question for the student…[they also] curate the questions a little and say 'Oh, a lot of students are asking about this.'" In this way, the TA was utilized in synchronous, formative assessment in order to ease the load from the professor during class in order to focus on delivering the lecture and providing answers to questions, rather than also having to monitor and consolidate student questions.

In another example, a professor cited having only one midterm and one final "in part because we're trying to reduce grading load for the TAs" and that much of the format of the exams strays away from short answer questions "because we have relatively new TAs [and] we're trying to give them experience grading, without overwhelming them." This professor emphasized the contractual obligation TAs were under to work only a certain amount of time per week, as well as acknowledging that many of their TAs were usually new to the position and they wanted to avoid "overwhelming them" with grading duties.

In both of these examples, TAs appear to be essential in carrying out assessment duties (particularly in large classes), whether those are formative in nature like the in-class question fielding, or summative in nature like grading exams. In either case, the voluntary and in-depth nature with which professors mentioned their TAs showed an appreciation for their support in

the classroom (particularly relative to assessment), as well as a recognition of their limitations (workload and experience).

While students typically mentioned TAs relative to their perceptions of motivation and power in the classroom, like professors, they too mentioned their involvement in classroom assessment as well. When asked about assessment experiences other than exams that one student had experienced, she cited "in discussions would be more of where the TAs are allowed to use…more creative forms of assessment." Students appeared to recognize that discussion sections (or smaller subsets of the overall class, usually 10-25 students each), was where assessment tended to vary more than traditional exams. These sections were usually administered by TAs, and thus, TAs were seen to be the face of the more non-traditional assessments.

Much like the earlier example in which the TA acted as the mouthpiece for students during class, it became clear that TAs could also be a source of empowerment:

> "I feel like [the] TA's a reflection of the professor and the TA's also are subject to the professor…so if the professor can give some leeway to that TA…I feel like the professor knows how to hand out some jurisdiction…so the professor doesn't take full control and full power…if the TA appears to have some sort of power…I feel safer in that classroom…because the professor is already giving out some power, maybe the professor is willing to hear [us] out."

This lengthy excerpt from a student illuminates many of the topics discussed with students and professors in one coherent monologue. While the TA here is clearly a go-between with students and professors, it appears TAs and their respective relationship with professors provides implicit curriculum to students about the extent to which students themselves will have power in that classroom. To the degree students perceive the professor being receptive to TAs and knowing "how to hand out some jurisdiction" or "leeway," to them, reflects whether students "feel safe" and have the potential to be "heard." In simple terms, when the TA appeared to have power in

their relationship with the professor, students felt they did, too. Thus, TAs appear to be an avenue through which perceptions of the classroom and attitudes towards assessment 'could be' administered by professors and communicated to students.

Unexpectedly, TAs emerged as an important middleman between students and professors: one that was not simply a passive operator, but rather, an important broker in communicating perceptions of the classroom and employing assessment practice.

***Positive Experiences.*** While for many participants the most salient experiences with assessment were those in which there were negative consequences or emotions associated, participants also spent time describing assessment experiences that were positive in nature.

As mentioned earlier, the *iClicker* polling system during class was one assessment method that was often cited by both students and professors. For both, instances in which this method was used to gauge understanding in the classroom were perceived as helpful and positive. In the example of one professor when asked about an assessment she found enjoyable, she cited the creation of *iClicker* questions as it "allows me to step outside of myself…and think 'Where are the places where the students might get caught?'" She goes on to mention "collaborating" with her LAs in order to get their "insight." Creating *iClicker* questions was a positive experience in that it allows her to evaluate her content delivery to pre-empt students' confusion which she finds "rewarding." Creating *iClicker* questions allows her to "step outside" herself in a possible reference to avoiding expert blind spot that can occur during instructional planning and consults her LAs in order to get their perspective on the accuracy of her estimations of student confusion.

Similarly, a student cited *iClicker* questions during class as one of her most enjoyable assessment experiences for the social aspect it provided. "There was this element of surprise

when you saw everyone else's response. It was almost like an emotional experience whenever we had questions, especially the ones where people really, really disagreed." Particularly for those moments in which students had variation in their *iClicker* responses, this student highlighted the "emotional" response of "surprise," that appeared to pique student interest. Capitalizing on the variance, the professor would then encourage a version of think-pair-share where students would then discuss with neighbors and resubmit *iClicker* responses after having convened. It is important to note that the instance that stuck out to her the most in this situation is not the *iClicker* questions in which she was right or where there was consensus in the class (much like the controlled motivation demonstrated by participants in 'Efficiency,'), but rather, those in which there were varying opinions and an opportunity to make those perspectives heard by crowdsourcing clarification on a topic. In the words of another student relative to think-pair-share: "[I] get to hear someone else's input…like 'Oh wow I saw it one way like, they saw another way,' and…[it] literally opens my understanding of things."

In all of these examples relative to *iClicker* questions or think-pair-share, it appeared the collaborative bent of these assessment practices was appealing for both students and professors—whether that was professors working with LAs/TAs to craft such assessments, or students working with their peers to evaluate and exchange ideas. These formative practices allowed for fluidity that appeared to stick with participants as positive and enjoyable.

In addition to *iClicker* questions and think-pair-share, professors and students also expressed positivity with novel assessments.

In one example, a student said she was asked to create a music video for a class project: "[It's] fun…it's not something you typically associate with organic chemistry…the things that we talked about in our music video, I still remember." This student makes the distinction

between “organic chemistry” and “fun”—such that those two are not concepts that typically overlap. While she describes the process as “hard,” she also discusses how that novel assessment format led to her remembering the content in the long run.

In another example, a professor describes a “manuscript peer review assignment” where students had to “wear a professional scientists’ hat” and make recommendations for a given empirical paper. She describes: “reading some of those has actually been really, really rewarding just to kind of, see ‘Wow! Students really picked up some good points here,’ or like, had some really nuanced ways of evaluating the empirical paper.” Instead of, perhaps, writing a paper to demonstrate understanding of scientific writing, the professor had students assume the role of journal editor in order to critique an empirical manuscript. To her surprise, it seems, this assessment allowed her to see the “nuanced” ways in which her students engaged in the assessment which she found a “rewarding” experience. In all, while sometimes an uncertain territory for both students and professors, novel assessments were met with positive reactions.

Finally, most positive assessment experiences reported by professors and students were those that related to real life. In describing an enjoyable assessment experience, a professor discussed an activity where students had to read and dissect a public health statistics table. She cited it as a “special instance where the students really see how statistics is useful in terms of like dispelling misinformation” and how the activity is “what I aim for as an instructor which is really connecting with their real life to the statistics.”

Similarly, another professor cites:

> “[I’m] interested in teaching my students valuable skills that they can take into their lives…because they're going to forget the facts—we all forget the facts…[we] memorize them for an exam [and] we forget…a writing assignment like an Op Ed [where] students choose a topic that's personally significant to them—I get feedback…that that's one of the more meaningful experiences for them…it's important to them”

In both of these examples, professors use novel forms of assessments—a critique of a public health statistic table, and an Opinion Editorial—courses in order to make content relevant to students' lives. They appear to recognize that these strategies allow for a "deeper" and "more meaningful" experience for students (which contrasts the unsaid 'shallow' and perhaps, 'less meaningful' experience of traditional forms of assessment). Rather than a focus on the minutiae of their course content, both professors here cite an emphasis on the "skills" from the course that can be carried into students' lives. This focus on a bigger picture appears to also expand the possibilities for assessing student learning as they are no longer confined to the smaller, more rigid details of course content.

Contrary to expectations, participants did not simply use our conversations to complain about the current state of assessment affairs. Participants related many positive experiences in their classroom assessment—from formative, in-class techniques like *iClicker*s or think-pair-share, to novel summative assessments that encouraged a focus on the larger scientific thinking skills as well as in students' own lives.

***Active Agent.*** Flexibility in assessment practice, TAs as the academic middleman, and overall positive assessment experiences lent themselves to students becoming and seeing themselves as active agents in the assessment process. In discussing his most enjoyable assessment experience, a student mentioned a group project he was able to develop in a lab section (overseen by TAs), where he was allowed to choose a topic of research:

> "I was able to pick my own topic…as opposed to [a topic] that gets researched by every other class, in every other quarter…I can create a study and…it's my study…[and] your name will be on the paper and this will represent who you are as opposed to filling in the blanks in [an] assignment."

In his example, the student demonstrates a strong sense of autonomous motivation. His frequent use of "I," "my," and "own," provide insight into how allowing students flexibility (even in the

way of topic choice), can be a source of power and, it seems also, pride. As opposed to something that is done in "every other class, in every other quarter," the student expresses his satisfaction with getting to create a study that is something he himself is able to "create," and have "represent" who he is. In this way, such assessment practices feel more personal to him in contrast to assessments that provide templates for students to "fill in the blanks."

This immersion of students as active agents in those assessments where they have choice, was also echoed by professors. In the example of the professor who asked students to write an Op Ed relating to course content, she cited the process as enjoyable for her, not only because they were "fun to read" given the varied topics and the way she was able to witness students "jump in with their perspectives," but also because of her students' reaction to it. She relayed: "[they] get so much more excited about doing that kind of writing than they do about writing a research paper." While the assignment was part of the course, the professor also provided a small dose of extra credit if students chose to polish their work and submit it for publication to a news outlet. Despite this portion not being required the professor found that many of her students opted to do so and had their work published in places like the "Daily Bruin" (UCLA's student newspaper). She found that not only were students "invested in creating…a quality final project…because they knew that it could actually be read right by other people" but they actually "wanted it to be read by other people."

In this example, students again, appear to be autonomously motivated by the ownership of the topic they chose—so much so that they were willing to spend the extra time to improve the products they submitted for a grade and go the extra mile to attempt to publish them. Their desire to have their work "read by other people," and create a "quality" product shows an active ownership of their work, and a desire to perfect it for an almost, intrinsic purpose. Being witness

to the process, the professor described this instance in a doting and glowing manner. In fact, this experience stuck out in direct contrast with the way another professor (as discussed earlier), mentioned that assessment was the "least favorite" part of her job. Here, the professor described the experience with such positivity that it imparted a sense of contentment and having achieved something that was mutually rewarding for both her and her students. Thus, students assuming this active agent role appeared to be something that professors enjoyed witnessing and being a part of.

Additionally, it appeared that students were 'activated' by non-traditional assessments that were challenging to them. For example, one student described the "flipped classroom style" in one class where material was learned outside of the classroom and class time was spent reviewing that material. The student described this practice as "great, because then the 'clicker' questions were harder and really like, tested if you understood what was going on or whether you needed more clarification." This stood in contrast to classes where *iClicker* questions could be answered "via context of the question" where there was "not really much substance to them. It's just like 'Oh—are you paying attention to the last thing I said?'…it's not as engaging." Ironically, those assessments that were used to gauge students' attention, were those that she found made her pay less attention. In her experience, having assessments that challenged her and made her aware of the gaps in her knowledge were more "engaging" and ones where she was more active, than those in which answers could be essentially guessed given "context" clues.

Similarly, another student said he appreciated when assessments "actually test to see if we understand the material deeply…to creatively apply a concept…in a way that you haven't seen before…like we were learning how to do something like in 3 dimensions and like tested to apply like, in 5 dimensions. I was like 'Woah!'" According to this student, being an active agent

in assessment meant having the opportunity in an exam or assessment where he could expand on his existing knowledge and demonstrate his ability to "creatively apply" what he learned. While many students discussed being able to engage "deeply" and be "creative" with non-traditional assessments, this student provides an example of ways this can be done, even in the context of a more traditional exam format.

Overall, students demonstrated fierce desire and enjoyment when they were active agents in their assessment and learning experiences. These experiences appeared to be equally enjoyable and satisfying for instructors as well. Choice, challenge, and change in the way assessment is typically employed appeared to encourage autonomy and autonomous motivation for students, as active, rather than passive (read, hostage), agents in their respective classrooms.

***Congruence.*** The way in which participants reflected what assessment "could be" almost always provided a sense of congruence. The most common way in which participants expressed congruence was between assessment and learning.

In one example, a student said they appreciated when their grading scheme showed varying percentages for assessments as opposed to 50% for a midterm and 50% for a final exam: "[When] 20% [was] allocated towards like, weekly quizzes or something where…you feel…your professor's actually like, making an effort to understand your progress in the…more holistic understanding of like, the class material." In their view, having varying percentages—particularly towards assessments that are formative in nature—would reflect the natural "progress" students make throughout the quarter, thus showing the "holistic understanding" students have of the material. In this way, the assessments would be congruent with the process of student learning: over time rather than discretely.

In attempts to decrease exam stress, a professor talked about the way in which they used quizzes "in a more low stakes environment," leading to "the less [students] complain and…[it] kind of decreases exam stress." By using lower stakes assessments, the professor is able to "align" or create congruence between what students practice with (quizzes) and their final summative assessment (the exam). The professor cites that this congruence appears to get across to students given their decreased complaints and stress relative to exams. In summary, a student cited "it feels good to be able to like, feel that you actually know the material, and then have the assessment reflect that, you know?"

This congruence was also mentioned in the use of flexibility in assessment and instances in which professors not only solicited feedback about the classroom and their assessments, but actually implemented some of the suggestions given. In the words of one student, such practice "really made [her] feel heard."

Relatedly, a professor cited that "UCLA with it being such a big school there's kind of a theme sometimes, of students feeling unseen…and so I try to do as much as possible to at least have them feel like, you know, they have a voice, it should be heard, and I wanted to hear it." This practice appeared to show congruence in instructor desire and what students perceived to be the purpose of such practice (feeling "heard.") Moreover, flexibility in the classroom allowed for the classroom (and potentially, too, its assessments) to congruently reflect student needs.

This same congruence was expressed for assessments and classrooms that were relevant to their lives. Whether, as mentioned earlier, students were allowed to choose their own project topic, or had assessments that in some way made their content "real" – like the plant presentations at the Botanical gardens—students demonstrated an appreciation for the way the assessments were congruent with their perceptions of the topic or their own lives. This was in

contrast with typical experiences where students expressed discontentment and an incongruence when heavy content was pared down to mere "true or false" questions, and exams that required "playing the game" in order to be successful.

A professor noted one facet besides grades in which she witnessed students be motivated. In her assignments she attempted to make them "inherently interesting" by using "popular" Press articles, podcasts, and 'Ted Talk' videos. She cited it as "rewarding" to hear that students were "forwarding the [sources] or talking about them to like family members or friends," because "class material became a topic…outside class discussions." The autonomous motivation described here mirrors that which students demonstrated when they described instances of relevant assessment experiences. Additionally, this excerpt goes back to the professor who claimed their "aim as an instructor" was to allow students to connect her content with their own lives. Such relevant assessment practices not only led to a perception of congruence on part of students, but also a perception of congruence for professors, in terms of their desire to have students view their content as not something simply confined to the walls of their classroom.

In this way, congruence of assessment with the learning process, having student voice sought out *and* used, as well as congruence between content and students' lives shows how assessment practices can align both with instructor purpose and student perception.

***Meaning & Purpose.*** A final outcome of assessment practices that demonstrate what "could be," was the meaning-making and purposes that participants described in their experiences heretofore. One such purpose, developing life skills, was that which participants hinted at throughout their experiences. For example, the professor who acknowledged the idea that students forget the "facts" when they leave the classroom, and therefore, she was more invested in imbibing "skills they can take into their lives": "I just don't know much of those

skills they can show on those exams and whether or not they carry it forward." In her view, the purpose of her classroom was to allow students to develop and retain life skills they could use beyond her classroom; thus, her assessments had to reflect such purpose by being non-traditional in nature (i.e., the Op Ed assignment).

Students, too, had similar conceptions of assessment purpose. For one student, this meant setting a personal goal to learn material by the *end* of the quarter. This was due to implicit curriculum of grading schemes which: "shows professors' expectations for how soon [we] should learn…if every midterm is worth the same…you're clearly saying that improvement doesn't matter…what you're saying is 'your performance in this one hour matters a lot'." In her meaning-making of assessments the student sees grading schemes as "professor' expectations" for the intended speed of their learning. As opposed to grading schemes with incremental increases in its grading percentage, typical grading schemes suggest that learning happens as discretely as they are numerated. In essence, it appears these percentages communicate to students the percentage of learning they should have 'achieved' at each assessment point.

In order to decrease stress when it came to summative assessments, one student noted that: "I think about the test as like, a celebration of my knowledge…and I think, thinking about tests in that way has really relieved a lot of my stress." In shifting her perception of tests, this student frames them as opportunities to "show" her knowledge. This shift in perception appears to help her decrease the pressure of high stakes assessments, but also hints at an unspoken norm. The fact that the student had to explicitly shift her perception of tests suggests that students do not naturally perceive "celebrating knowledge" as the purpose of tests, whether that be because of their formats, high-stakes nature, or other reasons.

In these two instances, students created their own meaning of assessments—whether that be determining the pacing of their own learning or reframing the purpose of a test—in order to, it appears, 'cope' with and alter the meaning that students typically derive from these assessments. While these students' ownership and creation of positive perceptions of assessments is indeed commendable, it does reveal the larger structural frameworks students operate within, and the heavy onus on individuals to adapt to their assessment environments, as opposed to vice versa.

Finally, one student summed the implicit curriculum and perception of meaning based on the type of assessments professors employed: "[If it's] a paper then it's probably because they want you to develop your own ideas versus…a multiple-choice exam…then they really want you to like know the content as is…without necessarily formulating your own ideas and arguments." To her, the format of an assessment indicated what the expected learning outcomes were for students, where a paper suggested wanting students to "develop" their own ideas, and exams suggested wanting students to show the "content as is." While professors likely develop their assessments with specific learning outcomes, it is unclear whether they realize the meaning students extract from these assessment choices such as grading scheme or assessment format.

Meaning and purpose of assessment practice appeared to go hand in hand with professors' reflection of their role in the classroom. In example of the professor who often used two-stage exams, he stated: "I increasingly view my role as an instructor as a facilitator of meaningful practice and feedback experiences and giving students opportunities to make errors and learn from them and be metacognitive." In another example, a professor describes the normal grading curve some of his colleagues use: "I just think they're terrible…it makes it look like our goal is to sort students and calibrate rather than to teach them." He went on to describe the extent to which he invites flexibility and student voice in assessment practice citing:

> "Students often will have suggestions for things they would rather do instead of what you want them to do and I'm usually very open to that because…my goal is not fairness because I'm not going to be sorting the students into a rank order. My goal is to help each student learn as much as they can."

Given his self-view as a teacher rather than "sorter," this professor is not opposed to flexibility and student voice in assessment as, from his perspective, teaching and learning relies less on objective "fairness," than does the role of someone attempting to "rank order" students. Thus, the professor cites being "open" to suggestions from students towards those assessment measures that will best serve the purpose of having students "learn as much as they can."

In both examples, the professors' view of their roles in the classroom—focused on student learning, being "facilitators" vs. "sorters," where mistakes are intentionally a part of the learning process—affected the types of assessments and openness to flexible assessments that involved student voice. In this way, the purpose of their assessments reflected this differing take on their roles as instructors.

Finally, students discussed the impact of choice in assessment on their meaning of assessment—more specifically, that of their motivation. One student described when offered choice: "I feel like I hold myself really responsible in mastery of the material…because *I* made the decision to be…assessed that way. Whereas…if I get a test and…I had no say in it…[it's] easier to put the blame on someone else rather than take responsibility for my own learning." In her experience, having the ability to have choice in assessment practice meant having to take responsibility for said choice. This autonomous motivation derived from the "active decisions," the student was able to make, in contrast to assessment experiences in which choice was nonexistent and placing "blame" was "easier" to do. Trusting students to make choices relative to assessment appeared to communicate meaning around students' capacity to take responsibility in their learning process, which was a motivating factor for students.

In all, it was clear that however participants experienced and described assessment experiences, assessments were in no way seen as objective, static tools used to measure learning in the classroom. Assessments—including the way they were evaluated, formatted, and viewed based on one's role and ability to make change—all carried with them certain underlying messages about the meaning and purpose around these assessments, which ultimately affected the way they were respectively perceived by each participant. These perceptions then affected various future actions (how professors decided to assess their students, or how students decided to engage with material), motivational reaction (autonomous vs. controlled), and what learning 'meant' in that context (i.e., memorizing vs. critical thinking).

**Classroom Observations and Syllabi Analysis**

Field notes from my classroom observations in addition to syllabi analysis of those classes, helped corroborate much of the experience participants recounted in our conversations. For example, the description of logistics such as classroom size where one professor described the "giant" classrooms where one feels like a "tiny fish" and a student felt like a "sardine," was also reflected in my own field notes of Site A (seating capacity 100-200) where I observed: "*large* classroom with *huge* screen made up of many smaller screens," and later "not hard to blend in with sea of students." The ocean metaphors noted separately by this professor, student, and myself, point to the striking vastness and physical logistics of many of these STEM classroom experiences.

The hostage experience was also communicated and summed in the syllabus for Site B where, for example, the heading for grading was bolded: "How We Will Assess Your Learning." The language here suggests an implicit divide between the "We" (teaching team) and the You(r) [students], whereby assessment is a one-way street from teaching team to students. In this way,

the students are given messages that the work of assessment is not for them and the judgement of their learning rests solely on the teaching team. Similarly, in the syllabus for Site A, the description of grading included being on an "absolute scale," with no room for change in final grades "in accordance with UCLA Academic Senate Regulation A-313 and strict rules governing the integrity of the grading process." The rigidity and literal "absoluteness" with which the grading in this course is described appear to preface how and why students may take on this hostage experience.

Finally, as mentioned by many participants, synchronous formative assessments such as *iClicker* opportunities were clearly a powerful strategy for transforming the classroom. In my field notes for Site B, I noted how tone of the classroom changed from "kind of dead," to "students sprung alive" with "more palpable energy" after a particularly divisive *iClicker* question that garnered lots of debate as to the right answer. In all, classroom observations and a deeper look at classroom syllabi helped to reveal much of the implicit curriculum participants described in their interviews with me.

A review of participant experience would not be complete without suggestions provided regarding how to turn "it is how it is" into what "it could be" relative to assessment practice at UCLA. Unlike previous sections, this one is specifically divided by student versus faculty suggestions in order to highlight the specific needs and wants of each population, as they are unique to each.

**Student Suggestions**

While the main aim of this study was to understand participants' current experiences with assessment practice, one underlying question was whether participants wanted change, and if so,

if they wanted to be a part of that change process. For students, the answer to this question was almost always yes.

To sum using the words of one student:

> "I definitely think students should be involved…these old standards of assessments can seem a bit out of date or archaic…students learn differently and…should be assessed in different ways…I would also want to understand the ways that I am assessing my own knowledge of the subject as opposed to just 'This is how it has always been, and this is how it's going to be."

Students expressed wanting their individual learning needs to be addressed by assessment practice, which was a factor in their desire to want to have their voice incorporated into classroom assessment. Moreover, as demonstrated by this student, students wanted to begin to develop that 'esoteric' expertise that has always been used to assess them, such that they could begin to accurately assess themselves.

So, what suggestions did students have for assessment practice and how did they envision their voice being represented in classroom assessment? Relative to this want for esoteric understanding of assessment, a student suggested professors provide "100 tricks and tips on how you study a question." The student appears to hint at wanting to eliminate construct-irrelevant variance such as test-taking skills/assessment literacy, in order to demonstrate his knowledge of the content better.

Another practice students mentioned across the board but explicitly mentioned as a suggestion from one student is: "please" use "clickers" in class. While the *iClicker* system does require specific technological equipment, students also mentioned similar polling systems that merely required internet access on any device, such as "Poll Everywhere" and "Mentimeter" (Warström & Ingvar, 2021). Relative to these polling features, students mentioned wanting

"harder" questions which appeared to refer to questions that probed more than simple comprehension or regurgitation.

One student suggested having a flexible range of dates for assignment submissions in order to account for the fact that "life happens." Generally, students asked for "more diverse ways to assess people," like including more quizzes (there were calls for both point-based and participation-based quizzes) as the quarter goes on, aligning such quizzes to the intended difficulty of exams, and helping students gauge their understanding at multiple time points.

Grading schemes that provided diversity in the sequence of percentage weight (to reflect the constructive process of learning), diversity in grading scheme types (where students could pick a scheme that worked to their strength), as well as diversity of assessment formats (quizzes, exams, papers, *iClicker*s, etc.) were also often suggested. Such practices would make students "feel like you're there to help them, more than see them fail."

One student mentioned the logistical obstacle of it being "impractical" for a professor to get to know each student and their respective assessment preferences, particularly in larger classes. Thus, she expressed want for more informal assessment: "Like, weekly reflections…even if it's just like, word vomit…[to] at least start thinking about the information." This harkens back to students preferring assessments that were relevant to them, where informal written reflections can serve that purpose while providing professors a glimpse into students' thought process without requiring formal grading.

This suggestion also echoes that of another student who said he would "prefer more writing and less testing," as well as group projects where one could "explain" their understanding of something "as opposed to just taking an exam and walking out the door." This student metaphorically suggests that students leave what they learned in an exam room when

they leave, whereas formats like writing and group projects allow him to take that knowledge with him beyond that room, that exam, and that class.

Finally, as mentioned in previous sections, students appreciated when faculty took time to ask for their feedback or how they thought something should be structured in the classroom; this was with the caveat that the professor act upon that feedback, so instructors who decide to adopt this practice should think about phrasing their feedback solicitation in a way that anything suggested within the stated confines are possible to act on.

In general, students typically jumped at the opportunity to suggest practices and ideas, which potentially hints at a lack of being able to do so in their assessment past. Moreover, students' suggestions were typically feasible in nature (as opposed to demanding a complete overhaul of current practices) and appeared to often acknowledge the limitations and logistics that professors fell under.

**Professor Suggestions**

Acknowledging the logistical constraints brought up by all participants, but particularly instructors, one of the interview questions asked professors to touch on both suggestions for future practice, as well as what they believe would be motivating in getting other instructors to adopt new ways of assessing and thinking about assessment.

Relative to suggestions for practice, one professor stood firm in his belief that two-stage exams were beneficial "to strike a balance between having challenging exams that are really good formative experiences…[where students] make a lot of errors on and learn from them." While the logistical challenge of exams may not be avoidable, the professor cites the importance of turning these into formative experiences for students such that they are encouraged to make

mistakes and learn from those mistakes. From his point of view, exams and formative learning experiences do not have to be mutually exclusive.

One professor suggested providing "alignment" such that professors "model for students the type of questions" that will be asked on assessments and provide "as much practice as possible" to "decrease exam stress." Similarly, another professor suggested: "you can get students to think about exams in a more positive light…[with] smaller assessments, making students think about assessment as feedback." While both of these professors suggest ways in to shift students' perception (from judgement to learning), they do not simply individualize the problem. Two-stage exams and modeling/practice assessments are meant to decompose the absoluteness of assessment consequences and the student perceptions of stress of certain summative practices. Relative to strategies that would encourage faculty to adopt flexible assessments that looked to include student voice, professors cited: ease, professional development/evidence, and student perception. In the words of one professor:

> "Realistically, you have to make it easy…it can't…require another 10 hours per week…faculty are motivated to publish…so just like students…'Well what do I get rewarded by, let me put my effort [there]…I think the same thing might be true of [faculty]."

Here, the professor stresses the importance of new assessment practice not adding a copious amount of work. Moreover, she empathizes with student efficiency and draws the comparison to the ways that faculty, too, look to put their efforts where it will be rewarded. This suggests that on a departmental level, promotion criteria would need to include innovation in teaching/assessment practices in order for professors to consider putting in the extra effort.

When consulting with CEILS, one professor appreciated: "just making it seem…like it's possible…it doesn't have to be a complete overhaul of what you're doing." Much like this professor, others too, mentioned the usefulness of resource offices such as CEILS and the Center

for Advancement of Teaching (CAT; UCLA: CAT, 2021) in helping digest and integrate such practices in their own classroom. These resource offices were especially helpful in the case of newer faculty, with one recounting coming into UCLA with "the message of, 'This is doable, we're going to help you and it doesn't have to be scary and dauting." These offices appeared to help faculty adopt practices with ease as well as establish a precedent for newer faculty who come into their positions with institutional messaging that such practices are encouraged and supported at UCLA.

Relative to existing faculty, one professor said it was crucial to "find ways to broaden [faculty's] definition of what counts as assessment," and perhaps "highlight" successful examples where student voice was incorporated in assessment practice. He went on to say: "most professors don't have any training at pedagogy…they care about students…but they don't …necessarily know what some of the options are." While not explicitly mentioned by the professor, it appears that CEILS and CAT could be useful in complementing professors' inherent "care" for their students and transform it into tangible practice.

One professor cited an institutional caveat that may also need addressing, namely "departmental norms," such that "giving students too much choice somehow violates those norms." Thus, he proposed a need to be "explicit" and "build consensus" among faculty. The professor touches on the unspoken "norms" within a department that may prevent faculty from feeling like they can extend such voice and choice to their students. In this way, being explicit in messaging and incentives within a department may aid in making these practices more common.

An additional way in which professors saw the potential for convincing faculty to engage in assessment practice that includes student voice is framing it from a diversity and inclusion lens. As one professor said: "The population at UCLA is so diverse…[and] practices from

students can be viewed as an effort to be more inclusive." For those faculty who seek to be inclusive in their teaching practices, highlighting the ways in which including student voice in assessment can cater to the diverse population at UCLA is seen as a way in which faculty may be persuaded to adopt such practices.

Finally, one professor cited a very personal way to encourage faculty to adopt such practices by suggesting an "appeal" to faculty as "lifelong learners" and their "sense of wanting to explore." For example: "'I bet when you were a student you would have loved to have an opportunity to do so and so,'" as a way to incite "retrospective evaluation." In his suggestion, inviting professors to think about their own love for learning might be a way to have them consider how they might foster that same trait in students by engaging in practices that were not necessarily granted to them in their educational experiences.

In general, professors cited many practices which suggest a shift to assessment practices that seek to be inclusive and incorporate student voice. Ease (including resource availability and department norms), evidence, student perception, and personal appeal were strategies that faculty claimed could be helpful in marketing these assessment strategies for a wider range of professors to adopt.

**Study 1 Conclusion**

Reflecting on the professor who described open ended assessments where students had choice but fixated on the "right answer," is an important representation of the state of classroom assessment currently. Having likely always had professors or teachers provide judgement on their work, students may not have had the chance to develop an internal gauge of their own work leading to this dependency on the professor as to what the "right answer" is (a reflection of "convergent" assessment practice; [Pryor & Torrance, 1997]). This incongruence highlights the

need to begin incorporating student voice into assessment practice such that students can develop their own sense of the "right answer."

Given the broad overview of "it is how it is," and suggestions for how "it could be," how do we conclude? Unfortunately, here too there is no one right answer. Participants' experiences touched on the complexity of classroom assessment practice, citing the diverse constraints, needs, purposes, and desires of students and professors. This study did not address *all* stakeholders involved in assessment practices, like TAs, departments as a whole, administration, and more—thus, one can imagine how much more complexity remains to be explored. That being said, participants shined light on some important possibilities for a classroom assessment future that can lead to more equitable, enjoyable, and feasible experiences for students and professors.

Referring back to the example one student provided of being offered the opportunity to choose test format, students' ultimate choice here is one that deserves a second look. With an opportunity to choose from a variety of test formats, students in this case chose, arguably, the most common (almost 'default') option of test format. In fact, most students and professors explicitly stated multiple-choice exams as the most common form of classroom assessment at UCLA. One question that arises then, is whether this choice was made as a result of familiarity with the format and the assessment histories that students bring to the classroom? That is, was this choice merely a result of the status quo experienced by students in most of their academic career?

A final thought here is whether the professor in this example anticipated student choice. Had the professor engaged in a similar poll of student preference before? And had this poll typically resulted in the same responses? If so, did the professor continue to engage in the

practice as a way to appear flexible in their assessment practice (whilst not actually having to alter practice)? That is, does the mere illusion of choice apply in this context, such that students' perception of the classroom can benefit from an almost rhetorical solicitation of their opinion? These are additional questions to which I do not have an answer and suggest empirical investigation in the future—particularly that of assessment histories and illusory choice.

I would like to conclude with the most surprising finding in this study that strongly suggests a need to incorporate students into classroom assessment practice. Of the three student participants who mentioned being a part of the LA program, all three expressed similar autonomous motivation such that grades and the assessment status quo no longer were a source of exerting effort or desire for certain outcomes. For this subset of students, their previous discontentment with classroom assessment produced an interest in learning about learning leading them to the LA program and the subsequent change in their personal views on the objective of learning in the classroom relative to assessment.

In an example of one of her duties, one LA student described working with an instructor to develop questions for an exam. She described the process as: "it made me feel like what I had to say really…mattered…my opinion was valuable…and that we're equals with these educators."

These perceptions held similar themes for the LA participants. While it is unclear whether their views are a result of the pedagogy training they undergo, the ability to work with professors on classroom assessments as "equals, or both—it is clear that students leave the program with positive perceptions of their classrooms overall. Providing insight into what an assessment future with students working alongside professors to develop assessments might look like, this student describes feeling like her "opinion was valuable" and that she felt a shift in typical classroom power dynamics such that she felt equal to both of the professors she had

worked with through the program. In general, the LA subset of participants was more likely to suggest feedback to their professors in the duration of their courses and have more autonomous motivation relative to assessments and learning. Finally, these students also tended to employ more understanding perceptions towards their professors and the way their classrooms were constructed—in acknowledgment of the difficulty of teaching and learning.

Again, while it is unclear what aspect of the LA program (or pre-existing attitudes) leads to such positive, and intrinsic perceptions of assessment and learning, this model of professors working with students to develop classroom structure (including that of assessments) certainly strengthens the ideas that learning about learning helped these students 'zoom out,' and consider their own broader philosophy about the work of the classroom.

**Link to Study 2**

Given these rich participant experiences of current assessment practices, the present dissertation sought to begin to envision the future of assessment practice with these experiences and suggestions in mind. By and large, most student participants expressed overwhelming desire to be involved in classroom assessment practice. Students touched on the want to develop their "assessor" identity, which heretofore appears to not have been given the space or frequent opportunity to develop. Additionally, students' enjoyment and eagerness to work with their peers in class showed an inclination towards active classroom participation. Thus, the evaluation of classroom participation materialized here as a potential avenue for students to be involved in the assessment dialogue. How might such a process be executed? As mentioned by many participants, TAs may be a key figure in bolstering perceptions of power and motivation for students, while carrying out assessment duties on part of professors that, to address many professors' concerns, does not require excessive time and effort on their part.

Many of the studies that look at student-centered classroom practices rarely focus on assessment, and those studies that examine the effects of student voice in assessment practice often do so qualitatively. Thus, there is a need for quantitative data that systematically points to possible effects of engaging students in assessment, as well as investigating the differential impact on students of various demographic identifications.

Combining factors uncovered in Study 1—students generally wanting to be involved in the assessment process, desire for more active participation and development of life skills, TAs as potential means of administering assessment practice that involves students, and professors' desire for empirical evidence that would such practice—leads to the conception of this second line of inquiry. Thus, the second study in this dissertation duo sought to engage student voice in the classroom assessment practice of participation evaluation—within discussion sections run by TAs—in order to empirically investigate the effects, if any, of such engagement on student perceptions of power, motivation, and attitudes towards assessment in the classroom of first year aspiring STEM students.

**Study #2**

The purpose of this study was to understand the effects of a classroom intervention that seeks to engage students in the assessment dialogue on perceptions of power, motivation… Engaging student voice in assessment practice may be an intimidating and confusing activity for students (Hewitt-Taylor, 2001). Additionally, while dated, there are concerns that students do not know enough about subject content coming into a course in order to be meaningfully involved in assessment practice (French et al., 1959). These concerns have been echoed by both professors and students in the data that has already been collected for Study 1. Thus, the current study

suggests involving student voice in an area of evaluation that does not necessarily require content knowledge: classroom participation.

In a survey of one urban university, 93% of all courses included participation as part of the overall course grade (Bean & Peterson, 1998). Relative to participation, a survey of 520 instructors in a large state university revealed 82% of faculty reportedly included participation in their syllabus, with only 25% of these professors actually providing criteria and grades for participation (Rogers, 2013). According to one professor as to why a grade was not formally assigned to participation, they noted: "I believe that different students learn in different ways and forcing quiet students to talk in class is obnoxious and likely to be counterproductive, e.g., superficial participation just for points" (p. 18). While this is a considerate perspective from the instructor's point of view, it actually serves to ignore the student voice of what good participation may look like (i.e., in the case of the "quiet student"), and rests on an esoteric assumption that good participation is represented solely by speaking up in class. By dismissing the evaluation altogether, the professor misses an opportunity to engage student voice in the co-creation of assessment criteria that may be representative of the diversity of students and their respective needs and preferences. This critical perspective that criteria must take into account minority students, cultural conflicts, and issues of representation and power in the classroom, relative to participation evaluation, has also been cited in the literature (Meyer & Hunt, 2011; White, 2011).

Given the prevalence of participation evaluation in course grades and the discord between its mention and its actual evaluation, I argue that participation is an appropriate realm to begin incorporating student voice in assessment practice.

The literature suggests some attempt to include students in the creation of participation criteria in the context of Australian higher education (Dancer & Kamvounias, 2005). In this intervention, students and tutors (instructors) submitted criteria they thought should represent participation in their course. The submitted criteria were reviewed by the researchers who made the final decision on criteria to be used for grading. Students were then allowed to self-assess their performance twice during the semester. Researchers analyzed for reliability of student assessment of participation versus their tutors and found that when criteria were clearly stated, reliability of participation was indeed high. While this study demonstrates students being somewhat involved in participation criteria development, it lacks the following: 1) the study failed to have students come up with the *operationalization* of participation skills (i.e., what behaviors are included under each criteria?), 2) the study did not measure the effects (if any) this practice had on students' perceptions of the classroom/instructor or psycho-social outcomes, and 3) the intervention and respective measurements were taken in a single semester with no longitudinal follow-up.

Thus, the current study seeks to include student voice in an intervention that allows co-creation of participation criteria in order to study its long-term effects on students' perceptions of power, motivation, and attitudes towards assessment in the classroom.

**Data Collection & Participants**

Participants in this study were undergraduate first-year students who enrolled in a STEM cluster course at UCLA in Fall 2020. The cluster program began as an initiative to aid in the college transition by creating "learning communities" focused within certain disciplinary topics where students take a series of courses for three consecutive quarters (one academic year). According to UCLA's web page on cluster courses, cluster students not only have higher GPA's upon graduation in comparison to their non-cluster taking peers, but also tend to graduate earlier

as well (UCLA Cluster Program, 2020). There is no official data relative to whether students maintain STEM major status or switch majors during this time. For this particular cluster, the grading scheme did not involve a curve. Moreover, the course did not serve the purpose of 'weeding' students out, but rather, fostering student interest in STEM fields.

In total, 240 first-year students were enrolled in the STEM cluster beginning in Fall, with some attrition during Winter (238) and Spring Quarter (232). Approximately 60% of participants self-identified as female and 40% as male. A third of participants identified ethnically as White, followed by 27% East/Southeast Asian, 14% South Asian, 16% Multiethnic, and 11% Latinx or Black/African American.

As the course took place during the COVID-19 global pandemic it was adapted for online instruction. In Fall and Winter quarters of the cluster course, students had access to pre-recorded lectures, alongside attending weekly synchronous Zoom discussion sections (with approximately 20 students per section). Participation in the discussion section comprised 10% of a students' total grade in the course. The weekly lecture was taught by the instructor of record (but also included guest lectures from other faculty), while the discussion sections were facilitated by graduate Teaching Assistants. It is here, within each individual discussion section, that the intervention was implemented.

The current study utilized an experimental, cluster randomization design to compare the effects of the intervention on perceptions of power, motivation, and attitudes towards assessment both between and within-groups. Intervention research in general often seeks to understand the effects of a treatment or intervention on one group with the effects of no such treatment/intervention on a comparable control group. This is an appropriate design for the

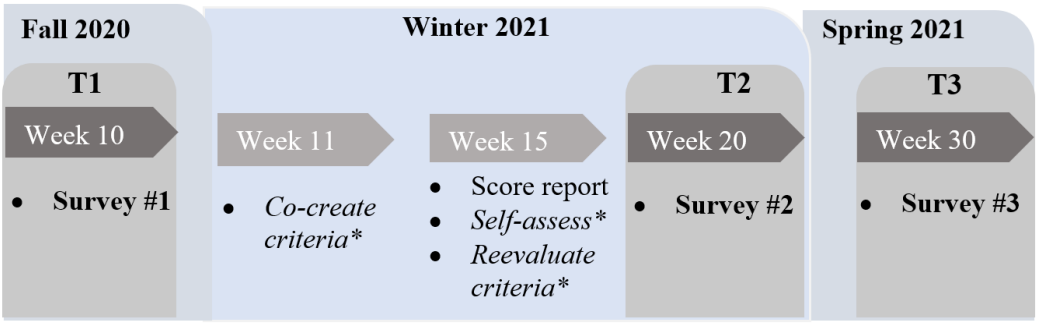
current study as it allows a comparison of students taking the same course, under similar conditions, both with and without the presence of the intervention.

An informed consent waiver was distributed to all students outlining their participation in the intervention in Fall 2020. Half of all discussion sections were then randomly selected to experience the intervention for the duration of Winter 2021. Teaching Assistants whose sections were randomly assigned to receive treatment in Winter attended a workshop at the end of Fall 2020 where the intervention protocol (outlined in the following section) was presented and standardized by the researcher such that all students experience the same treatment. The instructor was also present for this workshop.

During the workshop, the researcher carried out the intervention as though the TAs were students in the class (to allow TAs to understand the process from a student perspective). Then, TAs practiced creating grading progressions based on sample student criteria in order to calibrate a consistent standard across TAs. All materials required for the intervention (including a personalized script of intervention preface, Mentimeter poll, Google Docs [Google, 2021], etc.) were provided for each individual TA via a secured Google Drive shared only between the researcher and TA. This ensured materials were the same across the intervention, as well as allowed for 'process data' in order to ensure the intervention was carried out as intended. An email thread was also used between the researcher and intervention TAs and instructor in order to maintain uniformity across sections and answer any questions that arose about the process. Because the format of the course shifted from lecture plus discussion (in Fall and Winter) to pure discussion sections in Spring, the intervention only took place during Winter Quarter. Business as usual resumed for the Spring.

The first survey was administered at the end of Fall Quarter (T1) as a baseline of students' perceptions of power, motivation, and attitudes towards assessment, as well as key demographic information. This allowed time for students to acclimate and gauge classroom climate in order to respond to surveys accordingly. Following survey at T1, a second survey was administered at the end of Winter quarter (T2) in attempts to gauge any changes in these perceptions over time/as a result of the intervention being put into place. A final survey was administered at the end of Spring quarter (T3) in order to understand any lasting effects of the intervention from Winter quarter.

**Figure 2.1**
*Graphic Timeline of Intervention*




***The Intervention***

The overall aim of this intervention—as outlined in detail below—was to involve student voice in classroom assessment practice. More specifically, the intervention achieved the following: Firstly, it meaningfully engaged student voice in the assessment development process through the creation of participation evaluation criteria. Secondly, it allowed students an opportunity to stray away from the historical "dependence" (McCroskey & Richmond, 1983) on instructors for assessment evaluation, by allowing for self-assessment using the developed criteria. Additionally, as a result of having to create the criteria in addition to applying it via self-assessment, a final purpose of the intervention was to provide students a *holistic* experience—

from the very beginning of determination of purpose to the 'end result' of grading itself—of assessment in the classroom (generally solely experienced by instructors).

The proposed intervention took place on the first day of class in Winter quarter as the Teaching Assistant went over the policies of each section (see full protocol and sample products in Appendix C). To preface the intervention for students, the TA discussed the challenges of assessment in school contexts using the provided script. More specifically, the challenge of attempting to measure something, unlike weight or height, that is not tangible. The TA then cited that researchers, policy makers, professors, etc. have worked for decades trying to hone assessment practices to make them fair, valid for their outset purpose, and reliable. However, this iterative process has often failed to incorporate student voice in what is classified as important. The TA then expressed that in the current class, they want to give that opportunity to students such that they may co-create meaning of one aspect of assessment in the course: participation.

In guiding students to think about assessment purpose at large, the TA first asked the class as a whole, why participation may be a part of their grade (when it seemingly has nothing to do with STEM content). Probing questions included: "What might participation (in all its forms) be representative of? What skills might we be assessing when it comes to the various aspects of assessments?" If these questions proved to be too broad, the TA asked what other professors have articulated as good participation to them in the past, and why they think those specific things were considered important. For example, students might cite teamwork, communication, and respect as a few larger skills that are meant to be assessed by their participation grade. Based on this foundation, the TA then helped students develop the purpose of assessing participation in *their* class (e.g., what skills do "we" value and want to foster in this space).

Once a set of skills and the purposes of assessing participation had been established, the TA had students split up into breakout room groups and discuss what concrete behaviors might be representative of each of those skills (i.e., a skill of being respectful in the classroom might have a concrete behavior of not being distracted via cell phones or laptop use). Students noted these on a collective Google Doc shared with students, concluding with the TA facilitating a whole class consensus regarding the behaviors students come up with. Once developed, the TA noted that the criteria students co-created is what would be used to assess their participation. In order to scaffold the assessment development experience for students, a scale for grading was also suggested for the criteria heretofore created. Each discussion section offered a total of four participation points—one for attendance, and a maximum of an additional three based on the criteria set out. This scale was adapted from previous iterations of the course and was determined in conjunction with the professor of record. Three points were to be given for those students who exemplified the criteria above average, two points for average, and one point indicating a need for improvement.

Half-way through the quarter students were reminded of the criteria and had an opportunity to engage in self-assessment. To preface the quick write where they qualitatively self-assessed how they felt they lived up to the participation criteria they developed, students were provided copies of the criteria. Finally, students provided their own numerical score (0-3) of how they felt they had performed, on average, for the first half of the quarter. All sections (including non-intervention group) were provided their mid-quarter participation grades. This provided one way in which students in the intervention group could understand how criteria they developed resulted in their actual participation grade, but also provided an opportunity to show that the criteria they created were in fact being used by the TA. A Google Form (Google, 2021b)

was then sent to intervention students to ask if they wanted to update or change anything they initially came up with in the criteria, given their experiences in the first half of the quarter. No TAs reported any changes to the criteria they initially established.

## Operational Definitions & Measures

### *Power*

In the current study, power was operationalized as students' perceptions of autonomy support from their instructor in addition to their perception of having voice in the classroom. The 6-item "Learning Climate Questionnaire" (LCQ) [Williams & Deci, 1996] was adapted for the purpose of this study and was administered at T1-T3 (see Appendix D). Participants were prompted to "think about the way you are assessed by your TA and respond to the following prompts in regards to that assessment experience." Sample items include: "I feel that my TA provides me choices and options" and "My TA conveyed confidence in my ability to develop assessment criteria." Response options were on a 7-point Likert scale from 1-7 with 1 representing *Strongly Disagree* and 7 representing *Strongly Agree*. Item responses were aggregated into a single perception of power score for each participant ($a$=.88).

### *Motivation*

Motivation was operationalized here as approach/avoidance and mastery/performance orientation relative to this course. The "Achievement Goal Questionnaire-Revised" (AGQ-R) probing intersections of approach/avoidance and mastery/performance goals, often used with undergraduate populations, was administered at T1-T3 (Elliot & Murayama, 2008). Sample mastery approach items include: "My goal is to learn as much as possible" ($a$=.84). Sample performance avoidance items include: "My goal is to avoid performing poorly compared to others" ($a$=.85). Finally, sample performance approach items included: "My aim is to perform well compared to other students" ($a$=.81). A 5-point Likert scale was used for responses, with 1

being *Strongly Disagree* and 5 being *Strongly Agree* for each respective statement. As per validation findings for this measure as well as lack of operational clarity in the literature ((Elliot et al., 2011; Madjar et al., 2011), the mastery avoidance orientation was not included in analyses as it is not a significant predictor of intrinsic motivation nor actual performance.

### *Attitudes toward Assessment*

In the current study, student attitude toward assessment was operationalized as students' preference and beliefs regarding assessment in their classroom. A 5-item version adapted from the "Attitudes towards Grading System" scale developed by Pacharn, Bay, & Felton (2013) was used to gauge student attitudes. Sample items include: "I liked how the grading scheme employed in this course, with respect to participation, was determined" and "I believe that allowing students to participate in designing the grading scheme (e.g., in relation to participation) in a course wastes students' time that could be better spent working on the course material." Participants responded on a 7-point Likert scale with 1 indicating *Strongly Disagree* and 7 indicating *Strongly Agree*. Item responses were then aggregated into a single attitude towards assessment score for each participant.

### *Academic Achievement*

Final course grade percentages (which includes all course assessments from both lecture and discussion) served as a measure of students' academic achievement in this STEM course collected at each time point T1-T3.

### *Interest in STEM*

Three items probed student interest in STEM majors given their experience in the course, collected T1- T3. These included asking about students' comfort level with and belief about being successful in STEM, while the remaining asked about student inclination towards pursuing

a STEM major. A 7-point Likert scale was used for responses, with 1 being *Not true of me at all* and 7 being *Very true of me* for each respective statement (*a*=.77).

### *Covariates*

In addition to these measures, demographic information was surveyed. This included self-reported: age, ethnicity, gender identity, most recently attended high school, high school GPA, international/first-generation student status, parents' highest level of education as a proxy for SES, and any academic accommodations students received. Additionally, for the survey given at T1, students were asked whether they had any previous experience with choice and flexibility in assessment practice (*Yes* or *No*) in addition to the frequency (*Always, Very Often, Several Times, Once, Never*) and satisfaction of such experience (*Very Satisfied, Somewhat Satisfied, Neutral, Somewhat Dissatisfied, Very Dissatisfied*).

### *Qualitative Experiences*

For the survey administered in the intervention group at T2, a short answer section asked students to describe how the experience of being involved in assessment development made them feel, what effect it had on their perceptions of the classroom/instructor, what they enjoyed about the experience, and what might be used to improve the intervention. These questions provided qualitative data on students' experience of and suggestions to improve the intervention.

## Study 2 Results

### *Descriptive Statistics*

Corresponding means, standard deviations, and bivariate correlations of variables of interest are presented in Table 2.1 and 2.2.

The intervention group reported general declines in all motivational orientations, attitudes towards assessment, inclination towards STEM, and end-of-quarter grades from Fall to Spring.

Perceptions of power increased for this group from T1 to T3. Meanwhile, the control group reported declines in motivational orientations and end-of-quarter grades over time; perceptions of power, attitudes towards assessment, and STEM inclination generally increased for control participants over time.

For all students at T1, perception of power was positively correlated with end-of-quarter grade percentages ($r$=.248, $p$<.01) and attitudes towards assessment ($r$=.307, $p$<.01). Mastery approach was positively correlated with performance approach ($r$=.306, $p$<.01), performance avoidance ($r$=.186, $p$<.05), and attitudes towards assessment ($r$=.209, $p$<.05). Finally, performance approach was positively correlated with performance avoidance ($r$=.593, $p$<.01).

***Linear Mixed Models***

In order to answer the question of whether there were significant differences of key variables of interest within participants from Fall to Spring Quarter, as well as between the intervention and control groups, a random-slopes, linear mixed model was conducted in SPSS (Version 28; IBM Corp., 2017). Linear mixed models allow regression-like analysis on data that has a nested feature—in this case, students sampled from one class in their own individual discussion sections (UCLA: Statistical Consulting Group, 2021). This allows comparison of repeated measures longitudinally, without the assumption of compound symmetry (including covariance) [Magezi, 2015] and irrespective of missing data (UCLA: Statistical Consulting Group, 2021). The latter is especially pertinent to this study where not all participants were present on each data collection day ($n_{T1}$= 189 present, $n_{T2}$= 219, $n_{T3}$= 199) and those who were did not always complete every item during each collection point ($n_{T1}$= 44 incomplete, $n_{T2}$= 62 , $n_{T3}$=123). It should be noted that while there was a nested nature of participants in this study, this did not warrant the use of the multilevel command in the mixed model. This decision was made

based on recommendations by Paccagnella (2011) suggesting that level-2 variables should have a minimum of 50 units to accurately estimate error. In this case, the level-2 variable—discussion section—only totaled 12 pre-and during the intervention (Fall and Winter; two per TA) and 24 units post-intervention (Spring).

Seven distinct models were run: one for each of the outcomes of interest. Perceptions of power, attitudes towards assessment, STEM inclination, grades, performance approach, performance avoidance, and mastery approach goals each served as the dependent variable in their respective model (Tables 2.3 and 2.4). The model for each outcome of interest controlled for student ethnicity (White as reference), gender (Female as reference), and self-reported high school GPA. Predictors included the academic Quarter (T1-T3) and intervention group status. Participant ID was included as a random effect in order to account for within participant correlations. A sample model with power as the dependent variable is as follows:

$$Y(\text{Power})_{ij}=\beta 0 + \beta 1(\text{Quarter1})_{ij} + \beta 2(\text{Quarter2})_{ij} + \beta 3(\text{Intervention Group})_{ij} + \beta 4(\text{Ethnicity1-Latinx/Black})_{ij} + \beta 5(\text{Ethnicity2-Multiethnic})_{ij} + \beta 6(\text{Ethnicity3-E/S/SE Asian})_{ij} + \beta 7(\text{Gender})_{ij} + \beta 8\ (\text{HS GPA})_{ij} + \beta 9(\text{Quarter1 x Intervention Group1})_{ij} + \beta 10(\text{Quarter2 x Intervention Group1})_{ij} + u(\text{Participant ID})_{i} + \varepsilon_{ij}$$

Table 2.4 and 2.5 show main effects of the intervention and time on variables of interest. In all, there were no significant main effects of the intervention found for any outcomes. There were significant main effects of academic Quarter (time) on perceptions of power, quarter grades, and all motivation orientations of interest. Perceptions of power significantly increased for each subsequent time point, (standardized $\beta$ = 0.21, $p$=.018). All motivation orientations decreased from Fall to Spring. Mastery approach orientation decreased ($\beta$ = -0.07, $p$=.251). Performance avoidance decreased over time ($\beta$ = -0.11, $p$=.347) and performance approach also decreased from ($\beta$ = 0.07, $p$=.356). Finally, grades significantly decreased from Fall to Spring from an average of 99% to an average of 95.5% ($\beta$ = -0.02, $p$<.0001).

In order to understand the effects of the intervention on specific groups within the study, the following moderators were included as interaction terms in the above-described model: ethnicity, gender, prior choice in assessment, first generation status, and TA match from Fall to Winter (Tables 2.5 and 2.6). A sample interaction model for TA Match is as follows:

$$Y(\text{Power})_{ij}=\beta 0 + \beta 1(\text{Quarter1})_{ij} + \beta 2(\text{Quarter2})_{ij} + \beta 3(\text{Intervention Group})_{ij} + \beta 4(\text{TA Match})_{ij} + \beta 5(\text{Ethnicity1})_{ij} + \beta 6(\text{Ethnicity2})_{ij} + \beta 7(\text{Ethnicity3})_{ij} + \beta 8(\text{Gender})_{ij} + \beta 9\ (\text{HS GPA})_{ij} + \beta 10(\text{Group x TA Match})_{ij} + u(\text{Participant ID})_{I} + \varepsilon_{ij}$$

A significant interaction with intervention group and first-generation students was found for performance approach orientation (Figure 2.2). There was also a significant interaction of intervention group with whether TAs changed from Fall to Winter on perceptions of power (Figure 2.3). For those in the intervention group, there was a predicted .84 increase in first generation student performance approach orientation versus first generation students in the control group ($\beta$ = 0.84, $t$= 2.83, $p$ = .005). For those in the intervention whose TAs changed from Fall to Winter, there was a predicted .24 increase in reported perception of power ($\beta$ = 0.24, $t$= 2.28, $p$ = .024).

Given the non-significant main effect of the intervention, the final research question seeking to understand a moderated mediation effect of the intervention on performance via power and motivation was not tested. To sum, while the intervention did not have overall effects for all students in this context, there were moderator effects on perceptions of power for those who had a new TA during intervention implementation, as well as on performance approach orientations for first generation students.

### *Open Ended Responses*

***Intervention Group Students.*** In addition to gauging student experience and perception of the intervention with quantitative surveys, participants also had an opportunity to respond to

open-ended questions about their experience. Questions included how engaging in the process made them feel, how the process affected their perceptions of their classroom/instructor, and what suggestions participants had for improving the process. Responses were first filtered by whether the question was answered relative to the intervention itself (as prompted), or in regards to the class as a whole (omitted for these analyses). Sixty-nine participants answered at least one of the three prompts in relation to the intervention. EdWordle (Version 1; Wang et al., 2017) was used to create a word map of all the responses for each question and was then refined by filtering out filler words. In response to the first question of how the intervention made students feel, the following words were most commonly used: reflect/reflective (6), power/empowered (6), comfortable (3), control (4), heard (4), and included (2). In one student's words:

> "Although it was very short, I believe that it's a great technique to really establish that sense of learning within students. It places students at the center of their own success and achievement and that's really-really important for First Years and for students in general to be able to own up their own learning."

This was echoed in other responses that appreciated "having the autonomy to be able to implement [their] personal goals onto the grading criteria," and citing the experience as making them "feel very included and welcomed into UCLA."

While the majority of responses were positive, there were participants who cited neutral or contrasting stances to the experience. For example: "I think it was helpful to understand what we would be graded on, but it made me a little bit unsure about the grading at the same time since I am so used to teachers providing a grade just based on the amount of participation." In a similar vein, one student cited they "did not like it that much," and that "teachers should set the criteria and you should strive to meet those standards." Others found it "very nonchalant," and "unique" but not "particularly impactful."

In response to the second question of what effects engaging in the intervention had on students' perceptions of the classroom and/or their instructor, responses were again generally positive. One student responded that the classroom "felt more open and understanding, as more of a community rather than a prison." Another said the intervention showed the instructors as "accepting/trusting (treating us like adults haha)," while another said it showed the instructor "wasn't a tyrannical-stuck-up instructor." To sum for one student, the intervention helped show the classroom "as though I and the other students matter as people and have identities as such, rather than just as students. I felt that I could go to my instructor without judgement as well."

On the contrary, some students cited the intervention as having a "neutral" effect, while another echoed "It did not really have an effect. It did help me learn the material a little better though."

Finally, students were asked what could be improved about the intervention process (Table 2.7; *n*=28). A bulk of participants cited "N/A," "not sure," or something synonymous to "process was quite good" (*n*=13). One student simply said, "don't do it," while another thought "the process was structured perfectly." Suggestions for improvement included: having a more specific rubric of how each "subsection" of criteria mapped on to graded points, a reminder of the criteria more often throughout the quarter, and opportunity for "self-checks." While a preface of the activity was included in the script for implementation, one student said they would have liked "a little bit more guidance...especially for someone like me who had not done this process before so I was a little bit unsure about what the goal was." This was echoed in another comment with a student saying perhaps TAs can provide the "distinct categories of guidelines" (in this case, purposes of participation) and students could fill-in with criteria.

**Table 2.7**
*Intervention Improvement Suggestions from Students (n=28)*

- Post finalized criteria in publicly available space (i.e., CCLE)
- Quantify each subsection for grading purpose/clarity
- Allow participation self-checks/self-assessment
- Have participation grades available for viewing all Quarter
- More guidance in creating the criteria/provide guiding purposes
- More frequent check-ins about criteria and opportunity to adjust

***Intervention Group TAs.*** While students were the main focus of this intervention, TAs were also surveyed as to their experience implementing the intervention in order to understand the instructor perspective as well. The three intervention TAs all completed a short questionnaire at the end of Winter quarter about their experience conducting the intervention. The first question asked what TAs saw as some of the positive outcomes of co-creating criteria for participation with their students. They responded as it was a "good way to engage students" and made them "felt they were part of a community." Another said: "I think students are more relaxed about participation in that they don't feel like they have to be the most talkative one, and they feel more in control."

The second question asked about the challenges TAs perceived in co-creating criteria for participation. Only one TA responded to this question (the others: "N/A") explaining that this process was "more difficult than creating criteria myself because it requires facilitating a longer discussion."

The final question asked TAs how the experience of co-creating criteria with their students made them feel. One TA said it gave them a" better understanding of how the students experienced class - in terms of what they expected and wanted out of the experience - especially on Zoom," while another said "I like giving some of the authority and control to the students, as well as making the assessment more transparent. I think it's helpful for creating a supportive

learning environment." The last TA said they had already shared the idea with a community they were teaching with next quarter in attempts to "help build trust" with students.

**Discussion & Limitations**

The current study sought to longitudinally understand the effects of an intervention that engaged student voice in classroom assessment practice on perceptions of power, attitudes towards assessment, motivational orientation, STEM inclination, and academic performance. The significant main effect of time on students' perceptions of power in the classroom and motivational orientation point to the importance of studying student experience long-term, rather than cross-sectionally. Across the motivational orientations (mastery approach, performance approach, and performance avoidance), all students experienced a steady decline from Fall to Spring. This finding may point to the fatigue of the academic year (particularly in the fast-paced Quarter system), not to mention the toll of the global pandemic coupled with online learning. The gradual increase in student perceptions of power from Fall to Spring for all participants contrasts the motivational decline over time. This increase was perhaps a result of the consistent instructional staff that carried over from quarter to quarter (i.e., same professor and same group of TAs) which made it easier for students to have their voice heard.

To address the main interest here, the intervention did not have any significant main effects on any of the outcomes of interest. This was likely due to a couple of factors. For one, this course was far from what might be considered a "traditional" STEM course. Relative to assessment practices, the course did not curve grades (a very common occurrence in STEM classrooms in UCLA as cited by participants in Study 1), with the syllabus explicitly stating: "we do not grade this course on a curve, as research suggests that grading students relative to each other creates an unnecessarily competitive environment." Moreover, the very content of this STEM course was interdisciplinary. The course sought to view this particular STEM field

through the lens of "technical, political, cultural, and social dimensions." Thus, both the content and grading policies set this course apart from those that might be viewed as more typically rigid in nature (as was the case in the pilot implementation of the intervention [Chase, 2020]).

An additional reason why the intervention may not have had an 'above and beyond' effect is due to the bias in the sample available for this study. Finding instructor-collaborators in STEM for this work took three years; the instructor who became interested in the ideas presented in this study and was willing to allow their classroom to be open to this intervention, was one who was already quite invested in advancing equity through their pedagogical practices. In fact, they were recognized for their outstanding teaching in STEM in 2021 due to their concerted efforts in making their courses student-centered. Thus, a limitation here was availability of working with a 'traditional' STEM course/instructor. This is potentially because those who may not yet necessarily see the value in innovating their pedagogy were the same instructors who were not open to collaborating to incorporate this intervention into their course (and yet, may have had their course benefit the most given this intervention).

The intervention did, however, have modest significant effects for certain groups of students, although care must be taken in interpreting these findings given the number of tests. For those first-generation students in the intervention who reported an increase in performance approach as compared to their control peers, this finding suggests some motivational promise in incorporating student voice into assessment for those who are new to the nuances of higher education (and the assessment practices that accompany it). Performance approach has been shown to be important in the persistence and "bounce-back" for students who experience failure, thus an advantageous orientation to align with (Sideridis & Kaplan, 2011). These findings were similar for those intervention students whose TA changed from Fall to Winter and reported an

increase in perceptions of power in the classroom. While it was hypothesized that students with the same TA who experienced the intervention would experience significant increases in perceptions of power (due to familiarity with the TA), this finding suggests otherwise. It may be that when students encounter a new classroom with a new instructor (as is typically the case from quarter-to-quarter) this intervention may increase perceptions of power in that classroom space. Especially in this context where the content of the course stayed the same: the only difference was a new TA. These interaction findings demonstrate the potential stabilizing factor that the intervention may serve for students in new contexts.

All in all, while the intervention did not have a statistically positive effect for all, students' open-ended responses demonstrated a qualitatively positive experience. In the words of one student, the intervention made them feel: "kind of empowered. I felt heard and that my contribution mattered.; It made me feel like I need to take up more responsibility because we came up with these criteria ourselves, which I think is a good thing!" This comment points to the initial hypotheses during the conception of this dissertation, such that student voice in classroom assessment practice may motivate student achievement via perceptions of power and autonomy.

It is important to spend some time discussing a lurking set of conditions during data collection and intervention use: online learning plus the global pandemic. For these participants, this was likely the first college classroom experience at UCLA, that too, being exclusively online ('Zoom university') and physically disjointed from the UCLA community. Add to this the widening inequities exposed by the effects of the global pandemic (i.e., increased work and family responsibilities, particularly for already marginalized students, access issues with reliable technology, etc.). These conditions helped highlight the need for such intervention given the way the pandemic has forced instructors to rethink what was formerly taken for granted in

‘traditional’ classrooms. For example: How does one “participate” online? Do you physically raise your hand as you might in class? What if your camera or microphone fails to work? Or if your internet connection drops during the middle of a class discussion? The intervention helped clarify assessment criteria for participation *in an online format*, which was otherwise not something most had dealt with in higher education. In the words of one student: “I felt really supported which eased the online learning experience.”

Additionally, this intervention uncovered subtle inequity in current participation criteria for *in-person* classrooms. One student describes: “I enjoyed this because as someone with severe social anxiety it didn't make me feel pressured to be constantly speaking, in turn making me anxious about coming to discussion. It also made me feel like I matter and my opinion is in fact important.” This student points to the assumed participation criteria in in-person classrooms that synonymize participation with “constantly speaking.” The path for obtaining academic accommodations is strewn with barriers for students with disabilities (Toutain, 2019); thus, classroom assessment practice (including participation evaluation) may disadvantage those with ‘hidden’ disabilities or those who do not have formally requested accommodations. This points to yet another reason why student voice in classroom assessment practice is inevitably a necessity, towards the aim of creating more equitable classrooms.

I argue here that the forced switch to online instruction has led to a forced reflection of our classroom practices: including that of assessment. In this way, I encourage instructors to ask: What have we learned from the pandemic? How has the pandemic allowed us to question our norms about assessment and learning? And what lessons and practices do we want to keep moving forward? These questions help pave the way for future directions in the way of research and practice.

**Future Directions**

The first suggestion for future study in this area is implementing the intervention in the context of a more traditional STEM course. Moreover, while the current study supposed that first-year students would benefit most from having a say in assessment practice given their transition to college, this may be similarly beneficial to transfer students for whom UCLA is also just as new of a learning context.

Future research should also seek to understand what effects such intervention might have on other important psycho-social by-products of supporting student autonomy in the classroom such as self-efficacy, views of intelligence, and anxiety/stress which was often reported as a mental-emotional toll of current assessment practice. Additionally, given no significant correlation in this study between perceptions of power and motivational orientations, it may be useful to use a motivational measure which more closely aligns with autonomy and autonomous motivation (e.g., Motivated Strategies for Leaning Questionnaire (MSLQ), Pintrich, 1991) rather than performance/mastery and approach/avoidance measures which are more so treated as antecedents of intrinsic/autonomous motivation in the literature (Elliot et al., 2011).

Finally, it may be beneficial to explore this intervention with other marginalized student populations. While the current study did explore potential interactions with student ethnicity or gender, a larger sample size may be necessary to scope out these differences. As mentioned earlier, students with learning disabilities may benefit from having their voice heard in assessment practice. The current study did enquire as to whether students had classroom accommodations but given the barriers to receiving accommodations, many students with disabilities were likely left out of these analyses. Finally, this may also be useful work in the context of international students and English language learners whose voice may add a nuanced layer to assessment practice in the classroom.

The ultimate aim and suggestion here is for student voice to extend beyond participation to more meaningful content-based, assessment practice.

## Conclusion

Referencing the Critical lens through which this dissertation was envisioned, we must ask: *who* does our current assessment practice privilege? Which status quo does our current way of doing uphold? Themes from Study 1 suggest a disgruntled perception of "it is what it is" from both students and instructors. Findings from Study 2 suggest the need for more inclusive assessment practices that play to *all* students' strengths (in a variety of contexts), rather than to a select few.

To conclude, I would like to revisit Stiggins' (2018) call for revolutionizing assessment practice by "embracing a new role for students," and tie these dissertation studies together using the metaphor of the banyan tree as originally described by Jackson (2018). The "uncomfortable truth" of our current assessment practice can be demonstrated with a traditional tree: "as with assessment, the roots are out of sight, obscured, and people rarely comprehend how far they reach, and underestimate how much of the tree they comprise (p. 165-166)." In contrast, the banyan tree can "represent aspirations for fair and equitable ALT: here the essential sustaining root system of the tree is clearly displayed, and the paths for the nutrients from the soil to the leaves are made explicit and accountable."

I adopt this metaphor here with each student having their own 'tree.' Participants in Study 1 described being "hostage" to a system that was controlled by traditional assessment practice—demonstrating an "underestimation" of how important assessment practice (their roots) is to the overall workings of a classroom and aim of learning (each tree with its branches and leaves). For students in particular, being left out of the assessment conversation (being passively

watered and unable to see their roots) led to outcomes like “rebellion” (growing horizontally instead of vertically) and “efficiency” (using whatever means necessary to see quick growth, e.g., GMO’s, pesticide-use, etc.). What participants in Study 1 describe as “what could be” is the epitome of the banyan tree. Here, as in Study 2, students and instructors become “active agents” and co-gardeners of each tree. With roots now exposed, instructors and students can explicitly work together to find the combination of treatment that produces maximum growth for *each* tree; gardening (assessing for learning) is no longer one-size fits all. In this newly envisioned educational ecosystem, we seek to view the *trees from the forest*; and in doing so, foster a better growth environment for all.

**Table 1.1**
*Qualitative Interview Participant Demographics (Chronological Order)*

| Pseudonym | Title | Gender | Ethnicity | Discipline | Anticipated Graduation Year |
|---|---|---|---|---|---|
| Georgia | Student | Female | Indian/South Asian | Atmospheric and Oceanic Sciences + Humanities | 2021 |
| Vincent | Student | Male | Indian | Computational and Systems Biology | 2020 |
| Zach | Student | Male | South Asian | Biology-related switched to Humanities | 2020 |
| Susan | Student | Female | Indian | Psychobiology | 2022 |
| Renee | Student | Female | Indian | Physics | 2020 |
| Irene | Adjunct Professor | Female | Caucasian | Psychology + Another STEM field | |
| Natalie | Student | Female | South Asian/Punjabi | Psychology + Another STEM field | 2021 |
| Madeline | Assistant Adjunct Professor | Female | White | The Institute for Society and Genetics | |
| Annie | Assistant Professor | Female | Multiethnic | Psychology | |
| Bridget | Lecturer | Female | White | Psychology | |
| Julia | Assistant Professor | Female | White | Psychology | |
| Desiree | Student | Female | Indian/South Asian | Pre-Human Biology and Society | 2021 |
| Grant | Student | Male | Caucasian | Psychology | 2021 |
| Jerome | Professor | Male | White | Psychology | |
| Alison | Student | Female | South Asian | Psychobiology | 2020 |
| Sam | Student | Male | Caucasian | Psychology | 2021 |
| Cam | Student | Female | Southeast Asian | Psychology | 2021 |
| Shirley | Student | Female | White | Psychology | 2021 |
| Miranda | Lecturer with Potential Security of Employment (Assistant Teaching Professor) | Female | Caucasian | Psychology | |
| Joseph | Professor | Male | Middle Eastern | Physics & Astronomy | |
| Spiros | Professor | Male | White | Psychology | |
| Katerina | Student | Female | Multiethnic | Cognitive Science | 2021 |

## Appendix A:

## Semi-Structured Interview Protocol

**Student Version**

*Thank you so much for taking time to meet with me today. Your participation is invaluable to me and my research. Before we begin, I want to reiterate the confidential nature of this interview. Your name and any other identifiers will not be used in any way that might reveal your participation in this study. I also want to make it clear that your identifiers will not make their way back to your professors--the only kind of information that may be presented to them is about student experience in general, and about the ways they might improve their assessment methods (if they ask for such feedback). Please think of this as a conversation about your experiences, rather than a formal interview. I appreciate everything you would be willing to share, so long as you feel comfortable. With any research, there is always the possibility of feeling uncomfortable, and that is totally okay. If you feel uncomfortable at any point, you may always stop the recording, skip over sharing that which you feel uncomfortable with, or ask that we stop the interview. Do you have any questions before we begin?*

*Also, as a note about answering questions--its very helpful when you paint a picture for me and provide concrete examples about things that have made you feel a certain way, etc. so those details are very much welcomed.*

**Probing Anxieties about Assessment (without priming)**

13. When you enter a class on the first day, what are your biggest concerns or questions about the coming quarter?

**Establishing definitions**

*I want to clarify some terms I will be using in our conversation today. When we talk about educational assessment, classroom assessment, or assessment practice, I mean any of those strategies that have been used to evaluate your learning. This can include things like verbal questions from the professor during class, clicker questions, group work, exams, projects, papers, or portfolios. To be clear, when I say assessment, I don't just mean quizzes or tests. Does this make sense, or would you like some more clarification?*

**Situating assessment practice & Understanding what assessment experiences are important**

14. Given this definition, take a moment to think about all of your classroom assessment experiences here at UCLA. Can you describe a few of those moments that stick out the most in your memory?

    2a. Why do you think those instances in particular stuck out for you?

15. In thinking about the ways you have been assessed at UCLA, what are the most frequent ways your learning has been assessed?

3a. The least frequent?

16. What has been your most enjoyable assessment experience and why?

**Affective domain of assessment practice/experience**

17. When you think about being assessed in the classroom, how do you feel? What emotions do you associate with various assessment?
    5a. May need to break down into various types of assessment

**Typical concerns about assessment**

18. Imagine you are catching up with a classmate after class. What are some of the kinds of conversations would you be having about that class, particularly regarding assessment?

**Choice in existing assessment practice**

19. Have you ever been asked to participate in assessment practice? Whether that means coming up with the way you will be assessed, having a say in the assessment process, having flexibility or choice, etc. If so, describe that instance.
    7a. How did this experience make you feel?
    7b. Was this an experience you would want to engage in again?

**Personal definitions of power & assessment**

20. Switching gears just slightly, what does it mean to you to have power in the classroom?
    8a. What ways have you felt you have had power in the classroom before?
    8b. How does it feel to be in a classroom where you "have power" vs. where you feel you don't?
21. Can you describe the extent to which you feel you have power when it comes to classroom assessment?

**Motivation & assessment**

22. I want to get to know a little bit about motivation. Think about the type of assessments you have had in the past, and now think about your own motivation. How do these assessments factor into your motivation in the classroom, if at all?
    10a. Which assessments have served to increase your motivation? Why?
    10b. Which assessments have served to decrease your motivation? Why?

**Student Desire**

23. If your professors could anonymously get to know one thing about your assessment experience (i.e. something you would want to change, stay the same), what would that be and why?
    11a. In a similar vein, if you were given the chance to have a say in the way you were assessed in class, what would that look like? What would you suggest and why?

24. If you were asked to engage in assessment, is that something you would want to engage in? Why or why not?

**Feedback on Interview**

25. Were there any questions that you thought I might ask but did not?
26. Was there anything else you can think of related to this topic that you'd like to share?

*Thank you again for your time. This is all I have for today. I sincerely appreciate your willingness to be involved in this study, and keep in touch in regards to my findings. If you happen to know of anyone who might be willing to speak with me as well, please feel free to pass my contact information along to them so I may get a better picture of what the undergraduate experience with assessment is like here at UCLA.*

**In the case a participant gets overly emotional, clearly uncomfortable, stops the recording, or asks to terminate the interview:**

*Are you doing okay? Would you like to stop our conversation for today? I understand you'd like to pause our conversation today. Would you potentially be open to continuing this conversation at another time? No worries if not. As a note, if you need work through anything in our conversation that may have been too uncomfortable or triggering for you, we have the CAPS center at UCLA located next to the John Wooden Gym. They have walk-in hours until 4 or 5 PM everyday where you can consult with a fully licensed clinician. Would you like more information about this resource?*

**Professor Version**

*Thank you so much for taking time to meet with me today. Your participation is invaluable to me and my research. Before we begin, I want to reiterate the confidential nature of this interview. Your name and any other identifiers will not be used in any way that might reveal your participation in this study. I also want to make it clear that your identifiers will not make their way back to your students or departments--the only kind of information that may be presented to them is about professor experience in general, and about the ways assessment methods may be altered or the challenges that impede such alteration (if they ask for such feedback). Please think of this as a conversation about your experiences, rather than a formal interview. I appreciate everything you would be willing to share, so long as you feel comfortable. With any research, there is always the possibility of feeling uncomfortable, and that is totally okay. If you feel uncomfortable at any point, you may always stop the recording, skip over sharing that which you feel uncomfortable with, or ask that we stop the interview. Do you have any questions before we begin?*

**Context**

1. Can you describe the undergraduate course or courses you teach?
    a. How many students in each section?
    b. What are their average years? Transfer/international students?
    c. Is this a required course or one students take for GE's?

**Establishing definitions**

*I want to clarify some terms I will be using in our conversation today. When we talk about educational assessment, classroom assessment, or assessment practice, I mean any of those strategies that have been used to evaluate student learning. This can include things like verbal questions from the professor during class, clicker questions, group work, exams, projects, papers, or portfolios. To be clear, when I say assessment, I don't just mean quizzes or tests. Does this make sense, or would you like some more clarification?*

**Situating assessment practice & Understanding what assessment experiences are important**

1. Given this definition, take a moment to think about all of your classroom assessment experiences here at UCLA. Can you describe a few of those moments that stick out the most in your memory?

    2a. Why do you think those instances in particular stuck out for you?

2. In thinking about the ways you employed assessment at UCLA, what are the most frequent ways you have assessed learning?

    3a. The least frequent?

3. What has been your most enjoyable assessment experience and why?

**Typical concerns about assessment**

4. Imagine you are catching up with a colleague after a department meeting. What are some of the kinds of conversations would you be having about assessment?

**Choice in existing assessment practice**

5. Have you ever asked students to participate in assessment practice? If so, describe that instance.
    5a. How did this experience make you feel?
    5b. Was this an experience you would want to engage in again?
6. If you have not, what would you say have been reasons for you not doing so? (i.e., what are some challenges to implementing such practice?)

**Motivation & assessment**

7. I want to get to know a little bit about motivation. Think about the type of assessments you have implemented in the past. How do these assessments factor into your students' motivation in the classroom, if at all?
    7a. Which assessments have served to increase their motivation? Why?
        7b. Which assessments have served to decrease their motivation? Why?
8. From one professor to another, what do you think are strategies to engage professors in incorporating student voice into assessment practice?
    8a. What would motivate professors to want to engage in this kind of practice?

**Professor Desire**

9. If your departments could anonymously get to know one thing about your assessment experience (i.e. something you would want to change, stay the same), what would that be and why?
    11a. In a similar vein, if you were given the chance to involve students more in the assessment process, what would that look like? What would you suggest and why?

**Feedback on Interview**

10. Were there any questions that you thought I might ask but did not?
11. Was there anything else you can think of related to this topic that you'd like to share?

*Thank you again for your time. This is all I have for today. I sincerely appreciate your willingness to be involved in this study, and keep in touch in regards to my findings. If you happen to know of anyone who might be willing to speak with me as well, please feel free to pass my contact information along to them so I may get a better picture of what the professor experience with assessment is like here at UCLA.*

**In the case a participant gets overly emotional, clearly uncomfortable, stops the recording, or asks to terminate the interview:**

*Are you doing okay? Would you like to stop our conversation for today? I understand you'd like to pause our conversation today. Would you potentially be open to continuing this conversation at another time? No worries if not. As a note, if you need work through anything in our*

*conversation that may have been too uncomfortable or triggering for you, we have the CAPS center at UCLA located next to the John Wooden Gym. They have walk-in hours until 4 or 5 PM everyday where you can consult with a fully licensed clinician. Would you like more information about this resource?*

## Appendix B:

### Semi-Structured Classroom Observation Protocol

Describe the classroom:

- What does the physical classroom space look like?
- Describe resources, class configuration, #/ratio of professors, TAs/students, demographics, schedule, other important details

Describe the inner-workings of the classroom:

- What is the tone and tenor of the classroom?
- How do professors (and other adults) and students interact? Tone? Vocabulary? Purpose?
- What procedures (overt and tacit) exist? Who wields power and how is it negotiated? What happens in the classroom?

What happens in the classroom?:

- Describe the procedures, schedule, and overarching framework of the day.
- What is taught and how?
- Are students' interests and preferences for learning included? If so, how and to what degree?)

What happens in the classroom in regards to assessment?:

- How is student learning pre-assessed?
- How does the professor determine whether and to what degree the students have mastered the intended objectives during the lesson?
- How are discussions about assessment practice framed from the professor perspective? From the student perspective?

Who are the players in the classroom?:

- Describe the professor in action. What is their teaching style, her emphasis, her tone and emphasis?
- Describe the students, both collectively and individually. Write mini sketches of prominent (and prominent to you) students in the classroom.

# Appendix C:

## Sample Intervention TA Script & Protocol Packet

***Intervention: Part I***

**TA Preface Script**

"*Assessment is something we do in every course that can be particularly challenging; with classroom assessment we are attempting to measure something, unlike weight or height, that is not tangible. There is a broad history of researchers, policy makers, professors, etc. working to hone assessment practices to make them fair, valid for their outset purpose, and reliable. However, this iterative process has often failed to incorporate student voice in what is classified as important. In our class, we want to extend this opportunity to you all so we can co-create meaning of one aspect of assessment in this course: participation.*

*So let's think quite broadly here first: why participation may be a part of their grade (when it seemingly has nothing to do with STEM content). In other words: What might participation (in all its forms) be representative of? Or What skills might we be assessing when it comes to the various aspects of assessments? I'm going to drop the link for a Mentimeter poll in the chat, and I'd like you to respond reflecting on these questions I just posed. In essence, what do you all want to see as the purposes of participation assessment in our course?*"

**Mentimeter Link**

[*Link will be posted for your specific class by Winter quarter here]*

- You will copy and paste the above Mentimeter link in the chat for students to access.
- After students complete the Mentimeter, you will go to "**Present**" in the upper right hand corner of Mentimeter, and **Share** your screen with students in order to consolidate 3-5 purposes of participation evaluation in your course.

**Consolidate**

- Using the chat function or students speaking aloud, begin a consolidation of the many purposes students will have suggested in order to narrow them down to **3-5 main purposes**.
- Once your class has decided on the 3-5 purposes, you will assign students to **breakout rooms of 3-4 students** each.
- Before releasing them to breakout rooms, you will copy and paste the link of the Google Doc (find below) and note:

"Now that we have come up with the purposes of evaluating participation in this class, what behaviors or criteria do we want to be used to demonstrate each of these purposes? For example, given a purpose of *Verbal Communication we might say a criteria is speaking aloud during breakout rooms [<--provide an example relative to your class purposes here].*"

### Google Doc & Consensus

[*Link will be posted for your specific class by Winter quarter here]*

After about **5-10 minutes** (as you see a dwindling number of folks typing in the document), bring students back to the main room

- **Share** your screen with the Google Doc open and give students **2 minutes** to reflect on the criteria
- Ask: "*How do you all feel about these criteria? Are there any criteria you would like to add or remove*?"
- If there are no questions or concerns, have students indicate whether they would like to finalize the criteria
    - You can use the thumbs up reaction in Zoom, raise hand, Yes/No reaction, create a poll, etc.
- Once agreed upon you will note:

> *"Thank you for engaging in this process. This process mimics what instructors do when creating any classroom assessment: deciding on purpose and then identifying observable criteria to meet those purposes (i.e., when creating a rubric for an assignment). These are the criteria that will be used in our class to evaluate participation. I will create a polished version of these criteria and post it on CCLE/email it to everyone for their record."*

### Polish Criteria

- After class, use the following template to populate your course purposes and criteria. Save as **PDF** and email/post to CCLE for students to access ***and*** post in shared Google drive/paste here.

## Participation Criteria for Our Class

| ***Purpose 1 Here*** | ***Purpose 2 Here*** | ***Purpose 3 Here*** | ***Purpose 4 Here*** | ***Purpose 5 Here*** |
|---|---|---|---|---|
| • *Criteria Here* | • *Criteria Here* | • *Criteria Here* | • *Criteria Here* | • *Criteria Here* |

### Keeping Track

[*Google Sheets gradebook link will be posted here*]

- Use the above link for keeping track of student participation points throughout the quarter

***Intervention: Part II***

## Participation Revisit

[*Link to survey will be posted for your specific class by Week 5 Winter quarter here]*

Students will respond to the **survey** you send them (posted above)

- I will provide to you the criteria students suggest for changing/adding and you will a conduct a **5 minute consensus** like you did in Part I
    - For example, in the pilot study students brought up a concern about the breadth of criteria. I addressed this by assuring students that this was not a punitive system in that they had to participate in *all* criteria, but rather supposed to be an inclusive system that helped students play to their respective strengths relative to participation.
- You will **update student criteria as necessary** on the PDF and post again if any changes were made
- I will also provide student self-assessments of their participation in the first half of the quarter
    - You may use this to corroborate your own evaluation of students up until Week 5
- Finally, you will **share students' participation point total** at this point
    - I suggest copying the "ID #" column and "Total" column as a way to anonymously send scores to all students at once

# TA Training Activity

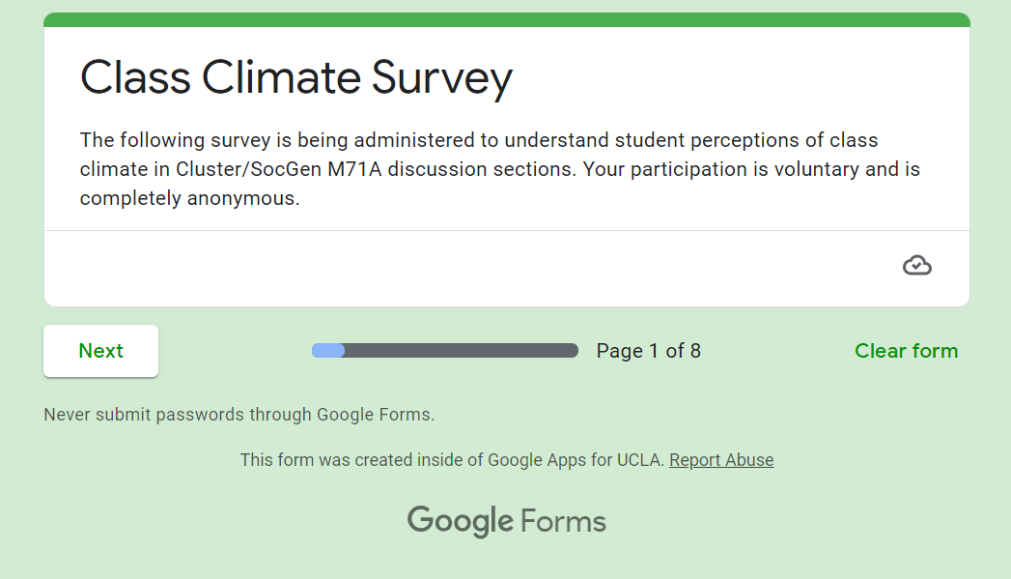
| Critical Thinking |
| --- |
| • Coming up with good ideas in breakout sessions/main room<br>• Expanding upon course material, including asking questions<br>• Relating responses back to what we are learning in lecture<br>• Sharing an original idea related to in-class materials<br>• Considering assignments in responses<br>• Sharing evidence from lecture and textbook (and potentially, real life) |

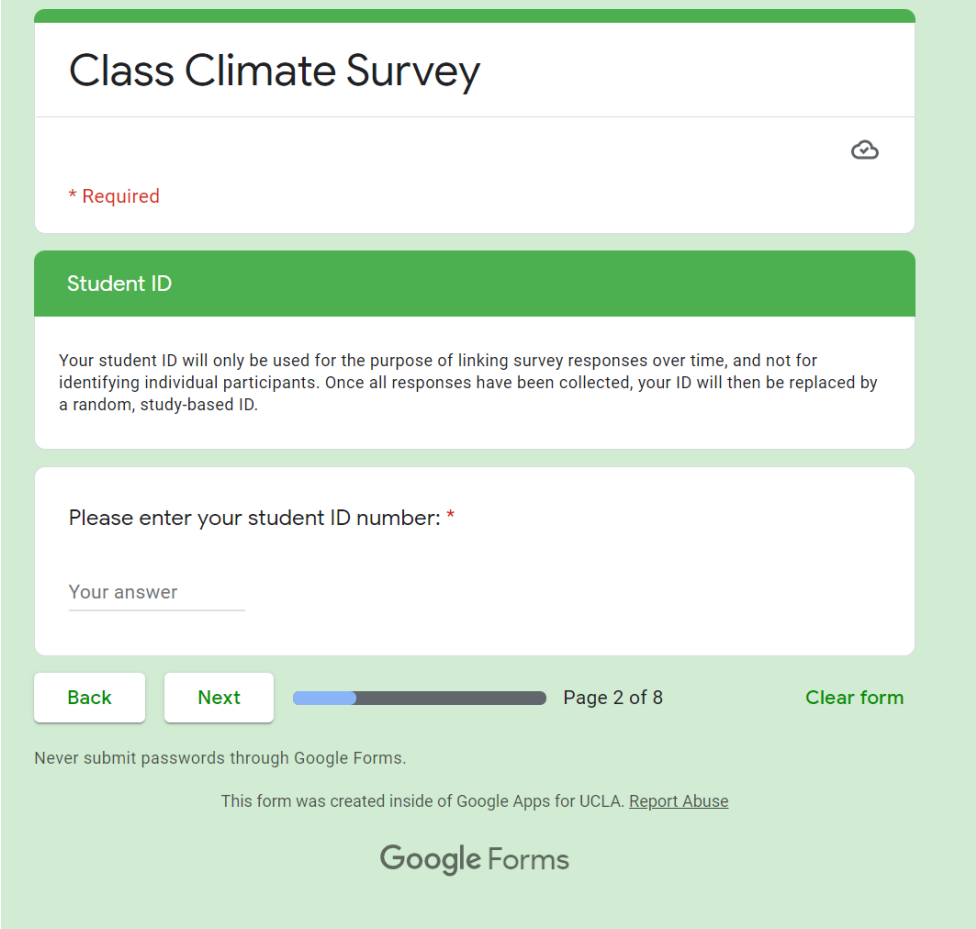
| Not Evident | Emerging | Developing | Controlled/Very Evident |
| --- | --- | --- | --- |
| • No clear use of criteria employed to suggest critical thinking | • Use of some criteria employed to suggest critical thinking<br>• Criteria demonstrated at a basic level with little purpose<br>• Criteria used sporadically | • Use of some criteria employed to critical thinking<br>• Criteria demonstrated at an average level with purpose<br>• Criteria demonstrated with some regularity | • Use of all or mostly all criteria employed to suggest critical thinking<br>• Criteria demonstrated at an advanced level with nuanced purpose<br>• Criteria demonstrated with regularity |

| Not Evident | Emerging | Developing | Controlled/Very Evident |
| --- | --- | --- | --- |
| Example: has not engaged in material or asked questions | Example: asks one question about material but is basic information from lecture/reading | Example: Relate response twice in a class period, to class material from throughout the quarter, tangentially | Example:Relate response twice in a class period, to class material from throughout the quarter, in a different light (adding new insight) |
| Example #2: | Example #2: | Example #2: | Example #2: |

## Appendix D:

## Sample Student Survey (Fall-T1)



# Class Climate Survey

## Part I

Think about the way you are assessed by your TA in discussion section, and respond to the following prompts in regards to that assessment experience:

1. I feel that my TA provides me choices and options.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

2. I feel understood by my TA.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

3. My TA conveyed confidence in my ability to develop assessment criteria.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

4. My TA encouraged me to ask questions.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

5. My TA listens to how I would like to do things.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

6. My TA tries to understand how I see things before suggesting a new way to do things.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

# Class Climate Survey

## Part II

Please respond to the following prompts.

1. My aim is to completely master the material presented in this class.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

2. I am striving to do well compared to other students.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

3. My goal is to learn as much as possible.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

4. My aim is to perform well relative to other students.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

5. My aim is to avoid learning less than I possibly could.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

6. My goal is to avoid performing poorly compared to others.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

7. I am striving to understand the content as thoroughly as possible.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

8. My goal is to perform better than the other students.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

9. My goal is to avoid learning less than it is possible to learn.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

10. I am striving to avoid performing worse than others.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

11. I am striving to avoid an incomplete understanding of the course material.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

12. My aim is to avoid doing worse than other students.

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

# Class Climate Survey

## Part III

Please respond to the following prompts:

1. I liked how the grading scheme employed in this course, with respect to participation, was determined.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

2. I believe that allowing a student to choose the criteria assigned to different components in their grading scheme (e.g., class participation) can help the student achieve a higher grade in the course.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

3. I believe that allowing a student to choose the criteria assigned to different components in their grading scheme (e.g., class participation) will likely increase the student's total work effort in the course.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

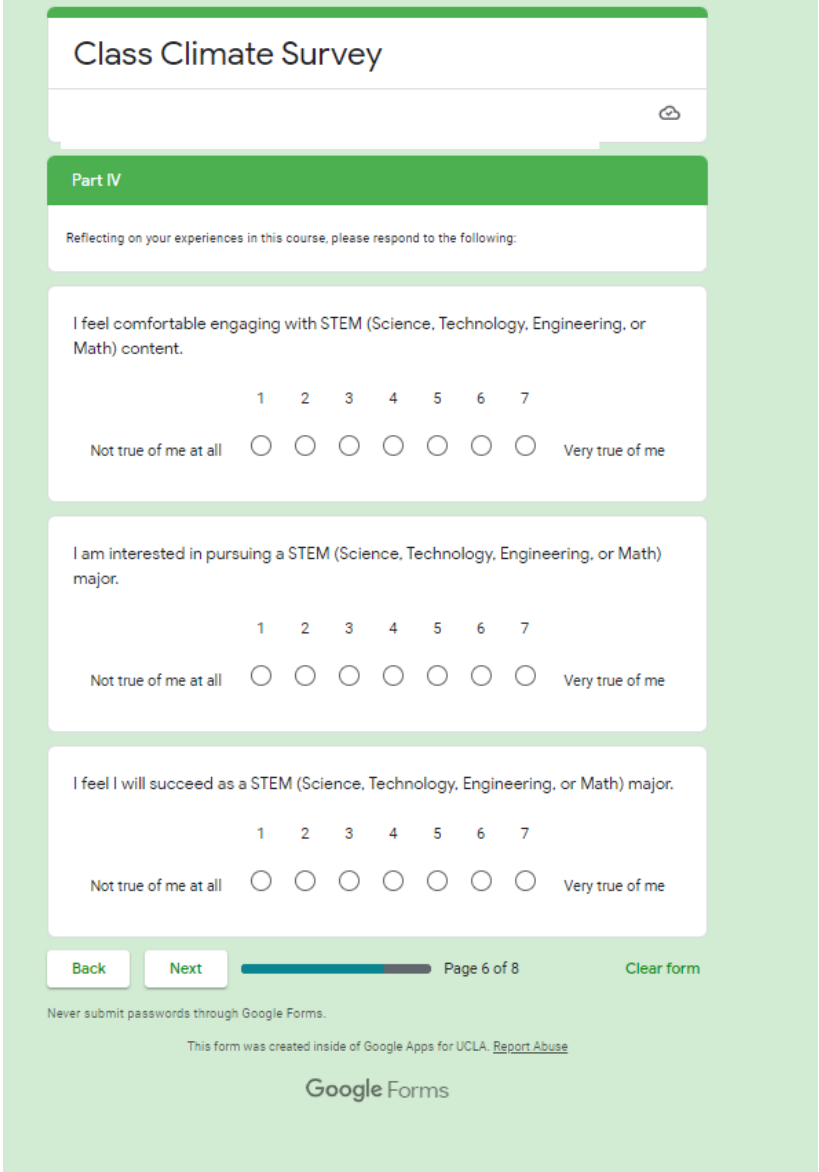
4. I believe that allowing students to participate in designing the grading scheme in a course wastes students' time that could be better spent working on the course material.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

5. If students are allowed to choose the criteria assigned to different components in their grading scheme (e.g., class participation), I believe they will be more likely to neglect some course activities that would be beneficial to them.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

Back Next Page 5 of 8 Clear form


Class Climate Survey
Part IV
Reflecting on your experiences in this course, please respond to the following:
I feel comfortable engaging with STEM (Science, Technology, Engineering, or Math) content.
1 2 3 4 5 6 7
Not true of me at all
Very true of me
I am interested in pursuing a STEM (Science, Technology, Engineering, or Math) major.
1 2 3 4 5 6 7
Not true of me at all
Very true of me
I feel I will succeed as a STEM (Science, Technology, Engineering, or Math) major.
1 2 3 4 5 6 7
Not true of me at all
Very true of me
Back
Next
Page 6 of 8
Clear form
Never submit passwords through Google Forms.
This form was created inside of Google Apps for UCLA. Report Abuse
Google Forms

# Class Climate Survey

## Part V

Please think about your classroom assessment experiences both in high school and here at UCLA (e.g., in class questions, tests, essays, group projects, etc.). Flexibility refers to any type of choice you were provided relative to classroom assessments. For example, choosing your grading percentages, dropping lowest scores, creating rubrics, etc.

Were you ever provided some sort of choice of flexibility in any of the assessments you experienced?

- ◯ Yes
- ◯ No
- ◯ Not sure

How often were you provided this choice or flexibility in any assessments you experienced?

- ◯ Always
- ◯ Very often
- ◯ Several times
- ◯ Once
- ◯ Never

If you experienced choice or flexibility in your assessment experience, how confident do you feel in the choices you made?

- ◯ Very confident
- ◯ Somewhat confident
- ◯ Neutral
- ◯ Somewhat unsure
- ◯ Very unsure

If you experienced choice or flexibility in your assessment experience, how satisfied do you feel with the choices you made?

- ◯ Very satisfied
- ◯ Somewhat satisfied
- ◯ Neutral
- ◯ Somewhat unsatisfied
- ◯ Very unsatisfied

Back Next Page 7 of 8 Clear form

# Class Climate Survey

## Demographics

High School GPA (weighted):

Your answer

Name your most recently attended high school:

Your answer

Please describe your ethnicity (check all that apply):

- [ ] Latinx/Hispanice
- [ ] Black/African American
- [ ] Native American/Alaska Native
- [ ] South Asian
- [ ] East Asian
- [ ] Middle Eastern
- [ ] White/Caucasian
- [ ] Native Hawaiian/ Pacific Islander
- [ ] Multiethnic
- [ ] Prefer to self-describe (enter in "other")
- [ ] Prefer not to answer
- [ ] Other:

What is your parent's highest level of education?

- ( ) Elementary/Junior high school
- ( ) High School
- ( ) Trade/Vocational School
- ( ) Community College
- ( ) College or University
- ( ) Graduate School
- ( ) Other:

Please describe your gender (check all that apply):

- [ ] Female
- [ ] Male
- [ ] Transgender Female
- [ ] Transgender Male
- [ ] Non-Conforming
- [ ] Prefer to self-identify (enter in "other)
- [ ] Prefer not to answer
- [ ] Other: _____

Are you an international student?

- ( ) Yes
- ( ) No

Are you a first-generation college student?

- ( ) Yes
- ( ) No

Please describe any accommodations you receive (check all that apply):

- [ ] None
- [ ] Technology-assisted notetaking
- [ ] Campus orientation and accessibility
- [ ] On-campus transportation
- [ ] Housing accommodations
- [ ] Proctoring and/or test-taking accommodations
- [ ] Alternate formats
- [ ] Sign-language/real-time capturing
- [ ] Prefer not to disclose
- [ ] Prefer to self-disclose (enter in "other")
- [ ] Other: _____

Age:

Your answer

Back Submit Page 8 of 8 Clear form

Never submit passwords through Google Forms.

This form was created inside of Google Apps for UCLA. Report Abuse

Google Forms

**Table 2.1**

*Summary of Variable Means and Standard Deviations over Time by Group (All: $n^{T1}=189$ $n^{T2}=219$, $n^{T3}=199$)*

| | All | | | | | | Intervention | | | | | | Control | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Fall (T1) | | Winter (T2) | | Spring (T3) | | T1 | | T2 | | T3 | | T1 | | T2 | | T3 | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Grade (%) | 99.32 | 5.49 | 96.58* | 6.61 | 95.76* | 4.44 | 99.73 | 4.62 | 97.55 | 4.70 | 96.16 | 3.67 | 1.00 | 3.92 | 97.87 | 3.41 | 96.22 | 3.41 |
| Power | 6.33 | .77 | 6.48* | .64 | 6.54* | .60 | 6.43 | .65 | 6.49 | .63 | 6.63 | .53 | 6.25 | .88 | 6.47 | .65 | 6.47 | .64 |
| Performance Approach | 4.08 | .85 | 3.97* | .84 | 3.72* | .94 | 4.17 | .82 | 3.95 | .83 | 3.87 | .85 | 4.06 | .87 | 3.98 | .85 | 3.55 | .99 |
| Performance Avoidance | 3.91 | .99 | 3.81 | 1.04 | 3.62 | 1.07 | 3.95 | .99 | 3.76 | 1.05 | 3.81 | .99 | 3.97 | .96 | 3.85 | 1.04 | 3.45 | 1.11 |
| Mastery Approach | 4.54 | .55 | 4.45* | .63 | 4.43* | .62 | 4.61 | .54 | 4.48 | .63 | 4.45 | .65 | 4.49 | .52 | 4.41 | .63 | 4.39 | .61 |
| Attitudes | 5.09 | .70 | 5.19 | .71 | 5.22 | .85 | 5.16 | .76 | 5.14 | .78 | 5.13 | .95 | 5.11 | .57 | 5.20 | .68 | 5.24 | .70 |
| STEM | 6.16 | .98 | 6.26 | .89 | 6.23 | .90 | 6.24 | .90 | 6.35 | .83 | 6.28 | .78 | 6.14 | .97 | 6.11 | 1.02 | 6.19 | .90 |

*$p$< .01, sig. change over time; Fall as reference

**Table 2.2**

*Summary of Bivariate Correlations for All Participants at T1 (n=189)*

| | Grades | Power | Performance Approach | Performance Avoidance | Mastery Approach | Attitudes | STEM |
|---|---|---|---|---|---|---|---|
| Grades | -- | -- | -- | -- | -- | -- | -- |
| Power | .25** | -- | -- | -- | -- | -- | -- |
| Performance Approach | .04 | .09 | -- | -- | -- | -- | -- |
| Performance Avoidance | .01 | .09 | .54** | -- | -- | -- | -- |
| Mastery Approach | .16 | .14 | .31** | .19* | -- | -- | -- |
| Attitudes | .03** | .31** | .15 | .14 | .21* | -- | -- |
| STEM | -.10 | .00 | .11 | .15 | .08 | .02 | -- |

*$p$<.05, **$p$<.01

**Table 2.3**

*Linear mixed model with Intervention Group Status and Longitudinal Effects predicting Perception and Performance Variables (n=195)*

| | Power | | | Attitudes | | | STEM | | | Grades | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | 95% CI | *p* | *B* | 95% CI | *p* | *B* | 95% CI | *p* | *B* | 95% CI | *p* |
| (Intercept) | 6.25 | 6.03 to 6.47 | .000 | 5.09 | 4.86 to 5.33 | .000 | 6.05 | 5.75 to 6.36 | .000 | 1.00 | .99 to 1.02 | .000 |
| Quarter[a] | .21 | .04 to .39 | .018 | .07 | -.13 to .27 | .487 | .03 | -.18 to .24 | .769 | -.02 | -.03 to -.01 | .000 |
| Winter | | | | | | | | | | | | |
| Spring | .23 | .03 to .44 | .027 | .16 | -.08 to .39 | .188 | .03 | -.22 to .28 | .827 | -.03 | -.04 to -.03 | .000 |
| Intervention Group[b] | .17 | -.04 to .38 | .118 | .05 | -.18 to .29 | .641 | .08 | -.21 to .27 | .585 | .00 | -.01 to .01 | .619 |
| Ethnicity[c] | -.06 | -.36 to .23 | .669 | .33 | .02 to .64 | .037 | -.19 | -.62 to .23 | .371 | -.03 | -.05 to -.01 | .002 |
| Latinx/Black | | | | | | | | | | | | |
| Multiethnic | .03 | -.22 to .29 | .795 | .01 | -.26 to .28 | .954 | -.02 | -.39 to .35 | .929 | .00 | -.01 to -.02 | .581 |
| E/S Asian | .02 | -.27 to .21 | .809 | .12 | -.08 to .32 | .257 | -.02 | -.29 to .26 | .907 | .01 | -.01 to .02 | .298 |
| Gender[d] | -.08 | -.25 to .09 | .361 | -.12 | -.29 to .06 | .191 | .22 | -.02 to .46 | .073 | -.01 | -.02 to .00 | .015 |
| HS GPA | .00 | -.01 to .01 | .344 | .00 | -.01 to .01 | .419 | .01 | -.01 to .02 | .465 | .00 | .00 to .00 | .428 |
| Quarter* | -.12 | -.37 to .13 | .342 | .00 | -.28 to .28 | .994 | .13 | -.16 to .43 | .383 | .00 | -.01 to .01 | .694 |
| Intervention Group | | | | | | | | | | | | |
| Winter*Int | | | | | | | | | | | | |
| Spring*Int | -.06 | -.35 to .24 | .711 | -.15 | -.48 to .17 | .353 | -.04 | -.39 to .32 | .832 | .00 | .00 to .01 | .838 |

[a]Fall=reference, [b]Intervention group=reference, [c]White/Caucasian/Middle Eastern = reference, [d]Female=reference

**Table 2.4**

*Linear mixed model with Intervention Group Status and Longitudinal Effects predicting Motivational Orientations (n=195)*

| | Mastery Approach | | | Performance Avoidance | | | Performance Approach | | |
|---|---|---|---|---|---|---|---|---|---|
| | *B* | 95% CI | *p* | *B* | 95% CI | *p* | *B* | 95% CI | *p* |
| (Intercept) | 4.72 | 4.52 to 4.92 | .000 | 4.12 | 3.77 to 4.46 | .000 | 4.27 | 3.70 to 4.57 | .000 |
| Quarter[a] | -.07 | -.18 to .05 | .251 | -.11 | -.33 to .12 | .347 | -.07 | -.23 to .08 | .356 |
| Winter | | | | | | | | | |
| Spring | -.09 | -.23 to .04 | .186 | -.38 | -.65 to -.11 | .006 | -.50 | -.69 to -.31 | .000 |
| Intervention Group[b] | .05 | -.13 to .23 | .597 | -.05 | -.37 to .27 | .769 | .07 | -.21 to .34 | .623 |
| Ethnicity[c] | .04 | -.25 to .32 | .795 | -.02 | -.46 to .51 | .926 | -.14 | -.58 to .29 | .507 |
| Latinx/Black | | | | | | | | | |
| Multiethnic | -.05 | -.29 to .20 | .710 | -.15 | -.58 to .27 | .475 | -.31 | -.68 to .07 | .108 |
| E/S Asian | -.12 | -.30 to .07 | .210 | -.22 | -.54 to .09 | .165 | -.27 | -.55 to .01 | .055 |
| Gender[d] | -.41 | -.57 to -.25 | .000 | -.12 | -.40 to .16 | .392 | -.17 | -.42 to .08 | .173 |
| HS GPA | .00 | .00 to .01 | .470 | .01 | -.01 to .03 | .304 | .01 | -.01 to .02 | .291 |
| Quarter*Intervention Group | -.04 | -.21 to .12 | .596 | .02 | -.30 to .33 | .919 | -.02 | -.25 to .20 | .831 |
| Winter*Int | | | | | | | | | |
| Spring*Int | -.10 | -.30 to .09 | .299 | .30 | -.08 to .68 | .119 | .16 | -.11 to .43 | .240 |

[a]Fall=reference, [b]Intervention group=reference, [c]White/Caucasian/Middle Eastern = reference, [d]Female=reference

**Table 2.5**

*Linear mixed model with Intervention Group and First Gen Status interaction predicting Performance Approach Orientation*

| | *B* | 95% CI | *p* |
|---|---|---|---|
| (Intercept) | 4.15 | 3.64 to 4.66 | .000 |
| Quarter[a] | -.09 | -.20 to .02 | .123 |
| Winter | | | |
| Spring | -.42 | -.55 to -.29 | .000 |
| Intervention Group[b] | .84 | .25 to 1.42 | .005 |
| Ethnicity[c] | -.32 | -.84 to .20 | .231 |
| Latinx/Black | | | |
| Multiethnic | -.36 | -.73 to .01 | .059 |
| E/S Asian | -.31 | -.59 to -.03 | .031 |
| Gender[d] | -.13 | -.37 to .11 | .297 |
| HS GPA | .01 | -.01 to .-02 | .210 |
| First Gen[e] | .16 | -.29 to -.61 | .484 |
| Intervention*First Gen | -.87 | -1.51 to -.24 | .007 |

[a]Fall=reference, [b]Intervention group=reference, [c]White/Caucasian/Middle Eastern=reference, [d]Female=reference, [e]First Gen students=reference

**Table 2.6**

*Linear mixed model with Intervention Group Status and TA Match interaction predicting Power*

| | *B* | 95% CI | *p* |
|---|---|---|---|
| (Intercept) | 6.20 | 5.97 to 6.42 | .000 |
| Quarter[a] | .15 | .03 to .27 | .016 |
| Winter | | | |
| Spring | .20 | .06 to .35 | .006 |
| Intervention Group[b] | .24 | .03 to .45 | .024 |
| Ethnicity[c] | -.07 | -.37 to .22 | .618 |
| Latinx/Black | | | |
| Multiethnic | .04 | -.21 to .30 | .751 |
| E/S Asian | .01 | -.18 to .20 | .940 |
| Gender[d] | -.05 | -.22 to .12 | .563 |
| HS GPA | .00 | -.01 to .01 | .401 |
| TA Match[e] | .19 | -.05 to .43 | .116 |
| Intervention*TA Match | -.35 | -.69 to .00 | .048 |

[a]Fall=reference, [b]Intervention group=reference, [c]White/Caucasian/Middle Eastern=reference, [d]Female=reference, [e]No TA Match=reference

**Figure 2.2**
*Differential effect of intervention for first generation students in intervention group vs. control on performance approach orientation*

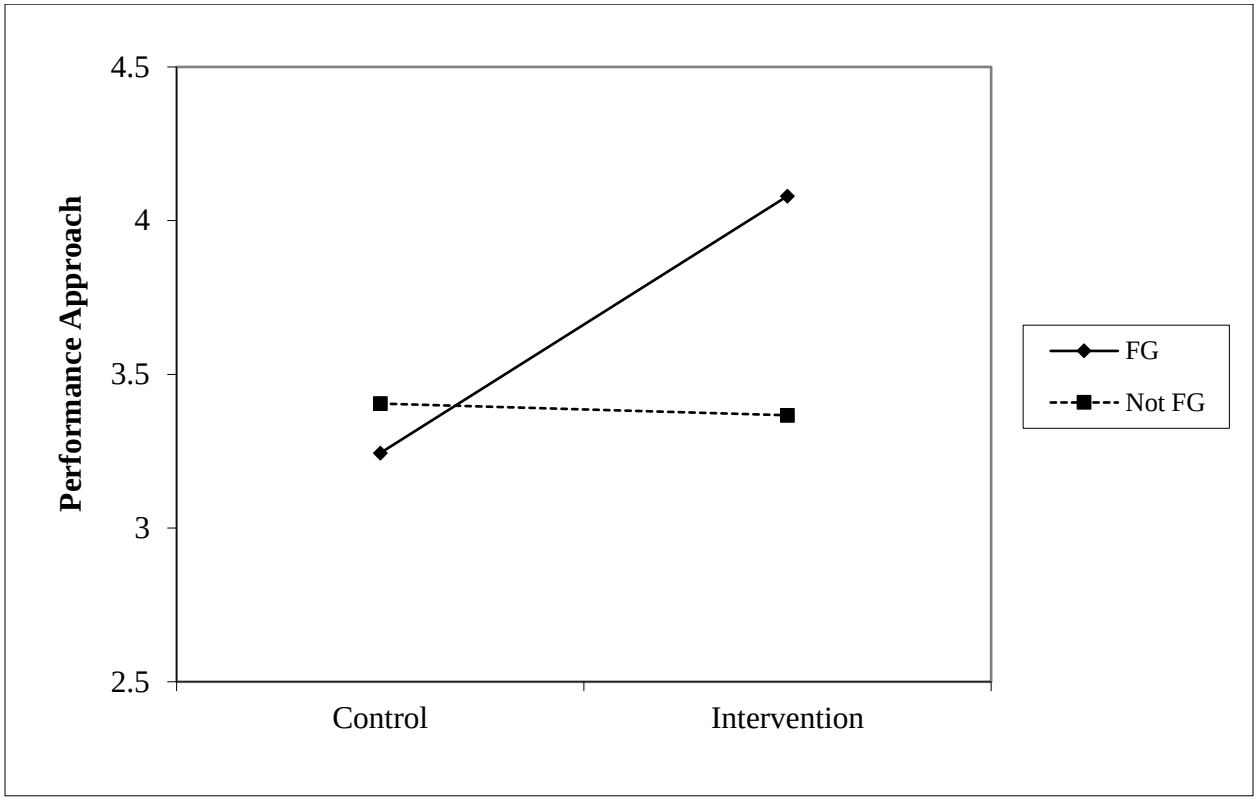
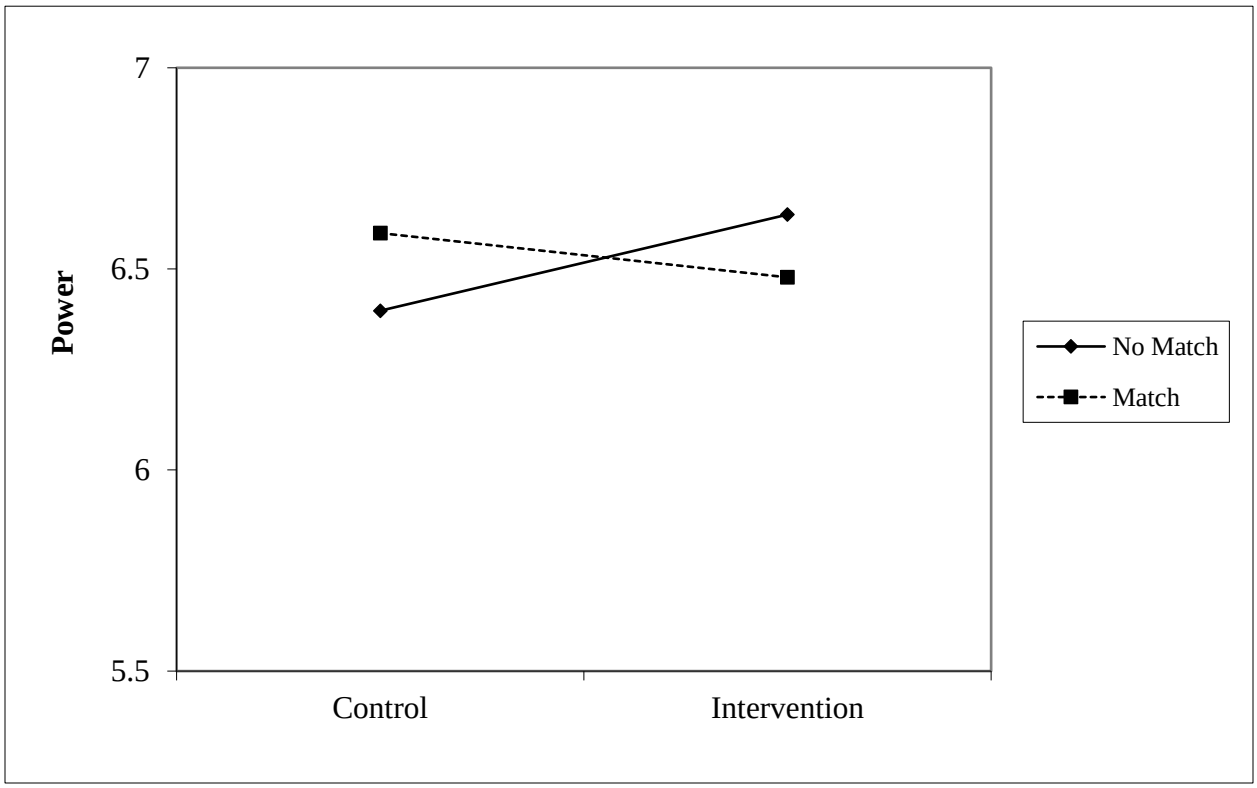


**Figure 2.3**
*Differential effect of intervention for students with no TA match versus TA match on perceptions of power*



References

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Asghar, M. (2012). The lived experience of formative assessment practice in a British university. *Journal of Further and Higher Education*, *36*(2), 205–223. https://doi.org/10.1080/0309877X.2011.606901

Ashenafi, M. M. (2017). Peer-assessment in higher education – twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education*, *42*(2), 226–251. https://doi.org/10.1080/02602938.2015.1100711

Astin, A. W. (1999). Student Involvement: A Developmental Theory for Higher Education. *Journal of College Student Development*, *40*(5), 12.

Bain, J. (2010). Integrating student voice: Assessment for empowerment. *Practitioner Research in Higher Education*, *4*(1), 14–29.

Bean, J. C., & Peterson, D. (1998). Grading Classroom Participation. *New Directions for Teaching and Learning*, *1998*(74), 33–40. https://doi.org/10.1002/tl.7403

Black, A. E., & Deci, E. (2000). The Effects of Instructors' Autonomy Support and Students' Autonomous Motivation on Learning Organic Chemistry: A Self-Determination Theory Perspective. *Science Education*, *84*(6), 740–756.

Black, P., & Wiliam, D. (2010). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan*, *92*(1), 81–90. https://doi.org/10.1177/003172171009200119

Bohman, J. (2016). Critical Theory. In *The Stanford Encyclopedia of Philosophy* (Fall Edition).

Bonner, S. (2013). Validity in Classroom Assessment: Purposes, Properties, and Principles. In *SAGE Handbook of Research on Classroom Assessment: SAGE Publications*. Sage.

Boud, D. (2007). Reframing Assessment as if Learning was Important. In *Rethinking Assessment for Higher Education: Learning for the Longer Term* (pp. 14–25). Routledge.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, *31*(4), 399–413. https://doi.org/10.1080/02602930600679050

Brighton, C., Moon, T., Jarvis, M., & Hockett, J. (2007). *Primary grade teachers' conceptions of giftedness and talent: A case-based investigation*. The National Center on the Gifted and Talented, University of Connecticut.

Broadfoot, P. (2007). *An introduction to assessment*. Continuum.

Brookhart, S. M. (2004). Assessment theory for college classrooms. *New Directions for Teaching and Learning*, *2004*(100), 5–14. https://doi.org/10.1002/tl.165

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980–1008. https://doi.org/10.1037/a0035661

Chase, M. K. (2020). Student Voice in STEM Classroom Assessment Practice: A Pilot Intervention. *Research & Practice in Assessment, 15*(2), 1–14.

Chen, X., & Soldner, M. (2013). *STEM Attrition: College Students' Paths Into and Out of STEM Fields* (p. 104). National Center for Education Statistics.

Chirkov, V. I. (2009). A cross-cultural analysis of autonomy in education: A self-determination theory perspective. *Theory and Research in Education, 7*(2), 253–262. https://doi.org/10.1177/1477878509104330

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891

Christopher Toutain. (2019). Barriers to Accommodations for Students with Disabilities in Higher Education: A Literature Review. *Journal of Postsecondary Education and Disability*, *32*(3), 297–310.

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, *7*, 205031211882292. https://doi.org/10.1177/2050312118822927

Cochran, S., & Spears, M. (1980). Student self-assessment and instructors' ratings: A comparison. *Journal of the American Dietetic Association*.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In *Self processes and development.* (pp. 43–77). Lawrence Erlbaum Associates, Inc.

Covington, M. V. (2000). Goal Theory, Motivation, and School Achievement: An Integrative Review. *Annual Review of Psychology*, *51*(1), 171–200. https://doi.org/10.1146/annurev.psych.51.1.171

Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education: Principles, Policy & Practice*, *3*(3), 265–286. https://doi.org/10.1080/0969594960030302

Daily, S., Eugene, W., & Prewitt, D. (2007). The development of social capital in engineering education to improve student retention. *Proceedings of the ASEE Southeast Section Conference.*

Dancer, D., & Kamvounias, P. (2005). Student involvement in assessment: A project designed to assess class participation fairly and reliably. *Assessment & Evaluation in Higher Education*, *30*(4), 445–454. https://doi.org/10.1080/02602930500099235

Deci, E., Connell, J. P., & Ryan, R. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, *74*(4).

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie Canadienne*, *49*(3), 182–185. https://doi.org/10.1037/a0012801

Deci, E. L., Vallerand, R., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and Education: The Self-Determinatio Perspective. *Educational Psycholgist*, *263*(3 & 4), 325–346.

Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

*Dedoose* (8.0.35). (2018). [Computer software]. SocioCultural Research Consultants, LLC.

Dweck, C. S. (1986). Motivational Processes Affecting Learning. *American Psychologist*, 9.

Dwyer, C. P. (2017). *Critical Thinking: Conceptual Perspectives and Practical Guidelines*. Cambridge University Press. https://doi.org/10.1017/9781316537411

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 × 2 achievement goal model. *Journal of Educational Psychology*, *103*(3), 632–648. https://doi.org/10.1037/a0023952

Evans, M., & Boucher, A. R. (2015). Optimizing the Power of Choice: Supporting Student Autonomy to Foster Motivation and Engagement in Learning. *Mind, Brain, and Education*, *9*(2), 87–91. https://doi.org/10.1111/mbe.12073

Fabert, N., Cabay, M., Rivers, M. B., Smith, M. L., & Bernstein, B. L. (2011). Exaggerating the Typical and Stereotyping the Differences: Isolation Experienced by Women in STEM Doctoral Programs. *American Society for Engineering Education*, 16.

Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Researc*, *59*(4), 36.

Fortier, M., Vallerand, R., & Guay, F. (1995). Academic Motivation and School Performance: Toward a Structural Model. *Contemporary Educational Psychology*, *20*, 257–274.

Foucalt, M. (1977). *The order of things* [Personal communication].

French, R., Raven, B., & Cartwright, D. (1959). The bases of social power. *Classics of Organization Theory*, *7*, 311–320.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research* (4. paperback printing). Aldine.

Google. (2021a). *Google Drive: Cloud Sotrage*. https://www.google.com/drive/

Google. (2021b). *Google Form: Online Form Creator*. https://www.google.com/forms/about/

Graebner, M. E., Martin, J. A., & Roundy, P. T. (2012). Qualitative data: Cooking without a recipe. *Strategic Organization*, *10*(3), 276–284. https://doi.org/10.1177/1476127012452821

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. In *Handbook of Educational Psychology* (Vol. 4, pp. 63–84).

Haas, N., Haladyna, T. M., & Nolen, S. (1990). *Standardized achievement testing: War stories from the trenches*. American Educational Research Association, Boston.

Haladyna, T. M., & Downing, S. M. (2005). Construct-Irrelevant Variance in High-Stakes Testing. *Educational Measurement: Issues and Practice*, *23*(1), 17–27. https://doi.org/10.1111/j.1745-3992.2004.tb00149.x

Harlen, W. (2005). *Teaching, learning and assessing science 5-12* (Vols. 5–12). Sage.

Hayenga, A. O., & Corpus, J. H. (2010). Profiles of intrinsic and extrinsic motivations: A person-centered approach to motivation and achievement in middle school. *Motivation and Emotion*, *34*(4), 371–383. https://doi.org/10.1007/s11031-010-9181-x

Hewitt-Taylor, J. (2001). Self-directed learning: Views of teachers and students. *Journal of Advanced Nursing*, *36*(4), 496–504. https://doi.org/10.1046/j.1365-2648.2001.02001.x

Horkheimer, M. (1982). *Critical Theory*. Seabury Press.

Hosek, A. M., & Houser, M. L. (Eds.). (2018). *Handbook of instructional communication: Rhetorical and relational perspectives* (Second edition). Routledge.

IBM Corp. (2017). *IBM SPSS Statistics for Windows* (Version 25) [Computer software]. IBM Corp.

*IClicker* (Version 7). (2020). [Computer software]. Macmillan Learning. https://www.iclicker.com/company/

Jackson, S. (2018). *Developing Transformative Spaces in Higher Education: Learning to Transgress* (1st ed.). Routledge.

Kinzie, J. (2020). How to Reorient Assessment and Accreditation in the Time of COVID-19 Disruption. *Assessment Update*, *32*(4), 4–5. https://doi.org/10.1002/au.30219

Knight, P. T., & Yorke, M. (2003). *Assessment, learning and employability*. Mc-Grah-Hill Education.

Koebler, J. (2012). Experts: "Weed Out" Classes Are Killing STEM Achievement. *U.S. News & World Report*. https://www.usnews.com/news/blogs/stem-education/2012/04/19/experts-weed-out-classes-are-killing-stem-achievement

Krippendorff, K. (1989). Content Analysis. In *International encyclopedia of communication* (Vol. 1, pp. 403–407). Oxford University Press.

Kusurkar, R. A., Ten Cate, Th. J., Vos, C. M. P., Westers, P., & Croiset, G. (2013). How motivation affects academic performance: A structural equation modelling analysis. *Advances in Health Sciences Education*, *18*(1), 57–69. https://doi.org/10.1007/s10459-012-9354-3

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an Enabler for Academic Success. *School Psychology Review*, *31*(3), 15.

Lofland, J., Anderson, L., & Lofland, L. (2006). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Wadsworth/Thomson Learning.

Lowery, G. (2010). Tougher grading is one reason for high STEM dropout rate. *Cornell Chronicle*. https://news.cornell.edu/stories/2010/04/tougher-grading-one-reason-high-stem-dropout-rate

Madjar, N., Kaplan, A., & Weinstock, M. (2011). Clarifying mastery-avoidance goals in high school: Distinguishing between intrapersonal and task-based standards of competence. *Contemporary Educational Psychology*, *36*(4), 268–279. https://doi.org/10.1016/j.cedpsych.2011.03.003

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, *6*, 2. https://doi.org/10.3389/fpsyg.2015.00002

Malcom, S., & Feder, M. (Eds.). (2016). *Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways* (p. 21739). National Academies Press. https://doi.org/10.17226/21739

Manganelli, S., Cavicchiolo, E., Mallia, L., Biasi, V., Lucidi, F., & Alivernini, F. (2019). The interplay between self-determined motivation, self-regulated cognitive strategies, and prior achievement in predicting academic performance. *Educational Psychology*, *39*(4), 470–488. https://doi.org/10.1080/01443410.2019.1572104

Martin-Hansen, L. (2018). Examining ways to meaningfully support students in STEM. *International Journal of STEM Education*, *5*(1), 53. https://doi.org/10.1186/s40594-018-0150-3

McCroskey, J. C., & Richmond, V. P. (1983). Power in the classroom I: Teacher and student perceptions. *Communication Education*, *32*(2), 175–184. https://doi.org/10.1080/03634528309378527

McGee, E. (2016). Devalued Black and Latino Racial Identities: A By-Product of STEM College Culture? *American Educational Research Journal*, *53*(6), 1626–1662.

McKellar, E. (2002). *Change our assessment practices? Why should we? The theory behind assessment practices.* Learning Communities and Assessment Cultures Conference, University of Northumbria.

McLean, H. (2018). This is the way to teach: Insights from academics and students about assessment that supports learning. *Assessment & Evaluation in Higher Education*, *43*(8), 1228–1240. https://doi.org/10.1080/02602938.2018.1446508

Menges, R. (1977). *The intentional teacher: Controller, manager, helper*. Brooks/Cole Pub. Co.

Merriam, S. (2009). *Qualitative research: A guide to design and implementation.* Jossey-Bass.

Messick, S. (1998). Test Validity: A Matter of Consequence. *Social Indicators Research*, *45*, 35–44.

Meyer, K. R., & Hunt, S. K. (2011). *Rethinking Evaluation Strategies for Student Participation*. *23*, 35.

Microsoft Office. (2015). *Microsoft PowerPoint* (Version 16) [Computer software].

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the Facts? Introductory Undergraduate Biology Courses Focus on Low-Level Cognitive Skills. *CBE—Life Sciences Education*, *9*(4), 435–440. https://doi.org/10.1187/cbe.10-01-0001

Naidu, S. (2017). How flexible is flexible learning, who is to decide and what are its implications? *Distance Education*, *38*(3), 269–272.

Naizer, G. L. (1997). Validity and Reliability Issues of Performance-Portfolio Assessment. *Action in Teacher Education*, *18*(4). https://doi.org/10.1080/01626620.1997.10463359

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, *14*(2), 149–170. https://doi.org/10.1080/09695940701478321

Nielsen, T. (2018). The intrinsic and extrinsic motivation subscales of the Motivated Strategies for Learning Questionnaire: A Rasch-based construct validity study. *Cogent Education*, *5*(1). https://doi.org/10.1080/2331186X.2018.1504485

Paccagnella, O. (2011). Sample Size and Accuracy of Estimates in Multilevel Models: New Simulation Results. *Methodology, 7*(3), 111–120. https://doi.org/10.1027/1614-2241/a000029

Pacharn, P., Bay, D., & Felton, S. (2013). The Impact of a Flexible Assessment System on Students' Motivation, Performance and Attitude. *Accounting Education*, *22*(2), 147–167. https://doi.org/10.1080/09639284.2013.765292

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, *28*, 151–160. https://doi.org/10.1016/j.concog.2014.06.007

Paris, S. G., Lawton, T., Turner, J., & Roth, J. (1991). A Developmental Perspective on Standardized Achievement Testing. *Educational Researcher*, *20*(5), 12–20.

Pintrich, P. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)* (Manual NCRIPTAL-91-B-004). National Center for Research to Improve Postsecondary Teaching and Learning.

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, *34*(1), 1–20. https://doi.org/10.1080/03054980701476386

Pryor, J., & Torrance, H. (1997). *Formative Assessment in the Classroom: Where Psychological Theory Meets Social Practice*. 26.

QAA. (2019). Quality Assurance Agency. *About Us*. https://www.qaa.ac.uk/about-us/what-we-do/our-work

Rogers, S. L. (2013). Calling the Question: Do College Instructors Actually Grade Participation? *College Teaching*, *61*(1), 11–22. https://doi.org/10.1080/87567555.2012.703974

Ross, J. A. (2006). The Reliability, Validity, and Utility of Self-Assessment. *Practical Assessment, Research & Evaluation*, *11*(10).

Ryan, R. M., & Niemiec, C. P. (2009). Self-determination theory in schools of education: Can an empirically supported framework also be critical and liberating? *Theory and Research in Education*, *7*(2), 263–272. https://doi.org/10.1177/1477878509104331

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed). SAGE.

Saulnier, B., Landry, J., Longenecker, Jr., H., & Wagner, T. (2008). Journal of Information Systems Education. *Journal of Information Systems Education*, *19*(2), 169–175.

Sharma, D., & Sharma, S. (2018). *Relationship between motivation and academic achievement*. *04*(01), 5.

Sideridis, G. D., & Kaplan, A. (2011). Achievement Goals and Persistence Across Tasks: The Roles of Failure and Success. *The Journal of Experimental Education*, *79*(4), 429–451. https://doi.org/10.1080/00220973.2010.539634

Sidky, G. (2017). The Power Game: Power Dynamics between the Teacher and the Students in a Graduate Seminar. *English Language Teaching*, *10*(5), 179. https://doi.org/10.5539/elt.v10n5p179

Simmons, A., & Page, M. (2010). Motivating Students through Power and Choice. *The English Journal*, *100*(1), 65–69.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571–581. https://doi.org/10.1037/0022-0663.85.4.571

Skjott Linneberg, M., & Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. *Qualitative Research Journal*, *19*(3), 259–270. https://doi.org/10.1108/QRJ-12-2018-0012

Smith, M. (2019). The science divide: Why do Latino and black students leave STEM majors at higher rates? *Washington Post*. https://www.washingtonpost.com/local/education/the-science-divide-why-do-latino-and-black-students-leave-stem-majors-at-higher-rates/2019/05/03/e386d318-4b32-11e9-93d0-64dbcf38ba41_story.html

Stiggins, R. (2014). *Revolutionize assessment: Empower students, inspire learning*. Corwin Press.

Stiggins, R. (2017). *The perfect assessment system*. ASCD.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.

Strauss, A. L., & Corbin, J. (1990). *Basics of Qualitative Research—Techniques and Procedures for Developing Grounded Theory* (1st ed.). Sage.

Suciu, L. (2019). Introductory Chapter: Discourse and Discourse Analysis. A Retrospective Approach. In L. Suciu (Ed.), *Advances in Discourse Analysis*. IntechOpen. https://doi.org/10.5772/intechopen.82823

Suddaby, R. (2006). From the Editors: What Grounded Theory is Not. *Academy of Management Journal*, *49*(4), 633–642. https://doi.org/10.5465/amj.2006.22083020

Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education*, *9*(2), 172–192. https://doi.org/10.1177/1469787408091655

Taras, M. (2012). Assessing Assessment Theories. *Online Educational Research Journal*, *3*(12), 1–13.

Taras, M., & Davies, M. S. (2013). Perceptions and realities in the functions and processes of assessment. *Active Learning in Higher Education*, *14*(1), 51–61. https://doi.org/10.1177/1469787412467128

Taras, M., & Davies, M. S. (2017). Assessment beliefs of higher education staff developers. *London Review of Education*, *15*(1), 126–140. https://doi.org/10.18546/LRE.15.1.11

Traub, R. E. (2005). Classical Test Theory in Historical Perspective. *Educational Measurement: Issues and Practice*, *16*(4), 8–14. https://doi.org/10.1111/j.1745-3992.1997.tb00603.x

Turner, S. L. (2014). Creating an Assessment-Centered Classroom: Five Essential Assessment Strategies to Support Middle Grades Student Learning and Achievement: This Article Presents a Classroom Assessment Strategy That Highlights Five Essential Elements of Instruction That Can Support Student Achievement. *Middle School Journal*, *45*(5), 3–16. https://doi.org/10.1080/00940771.2014.11461895

UCLA. (2021). *UCLA: Center for Education Innovation & Learning in the Sciences*. https://ceils.ucla.edu/

UCLA: Center for the Advancement of Teaching. (2021). *UCLA: Center for the Advancement of Teaching*. https://teaching.ucla.edu/

*UCLA Cluster Program*. (2020). UCLA: Undergaduate Education Initiatives. http://www.uei.ucla.edu/academic-programs/ucla-cluster-program/

UCLA: Statistical Consulting Group. (n.d.). *Introduction to Linear Mixed Models*. Institute for Digital Research & Education Statistical Consulting. Retrieved June 16, 2021, from https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/

UCLA: Statistical Consulting Group. (2021). *SPSS Mixed Command*. Institute for Digital Research & Education Statistical Consulting. https://stats.idre.ucla.edu/spss/seminars/spss-mixed-command/

van der Linden, W., & Hambleton, R. (Eds.). (2013). *Handbook of Modern Item Response Theory*. Springer Science & Business Media.

Van Soom, C., & Donche, V. (2014). Profiling First-Year Students in STEM Programs Based on Autonomous Motivation and Academic Self-Concept and Relationship with Academic Achievement. *PLoS ONE*, *9*(11), e112489. https://doi.org/10.1371/journal.pone.0112489

Wang, Y., Chu, X., Bao, C., Lifeng, Z., Deussen, O., Chen, B., & Seldmai, M. (2017). *EdWordle* (Version 1) [Computer software].

Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: The crucial factors of design, teacher participation and feedback. *Assessment & Evaluation in Higher Education*, *43*(7), 1032–1047. https://doi.org/10.1080/02602938.2018.1427698

Warström, J., & Ingvar, N. (2021). *Mentimeter* (3.2.3) [Computer software]. https://www.mentimeter.com/

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J., & Blakemore, S.-J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271. https://doi.org/10.1016/j.concog.2013.01.004

White, J. W. (2011). Resistance to Classroom Participation: Minority Students, Academic Discourse, Cultural Conflicts, and Issues of Representation in Whole Class Discussions.

*Journal of Language, Identity & Education*, *10*(4), 250–265. https://doi.org/10.1080/15348458.2011.598128

Williams, G. C., & Deci, E. L. (1996). Internalization of Biopsychosocial Values by Medical Students: A Test of Self-Determination Theory. *Journal of Personality and Social Psychology*, *70*(4).

Winget, M. A. (2008). A methodology and model for studying boundary objects, annotations, and collaborative practices: Musicians and musical scores. *Proceedings of the American Society for Information Science and Technology*, *44*(1), 1–13. https://doi.org/10.1002/meet.1450440238

Wise, S. L. (2020). The Impact of Test-Taking Disengagement on Item Content Representation. *Applied Measurement in Education*, *33*(2), 83–94. https://doi.org/10.1080/08957347.2020.1732386

Zoom Video Communications, Inc. (2020). [Computer software] (4.6.9).